621.396(06)
Р15

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ,
МОЛОДЕЖИ И СПОРТА УКРАИНЫ

ХАРЬКОВСКИЙ НАЦИОНАЛЬНЫЙ
УНИВЕРСИТЕТ РАДИОЭЛЕКТРОНИКИ

РАДИОТЕХНИКА

Всеукраинский межведомственный научно-технический сборник

Тематический выпуск

ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ

Основан в 1965 г.

ВЫПУСК 171

Радиотехника
621.396(06) Р15
827248
НБ ХНУРЕ 2012

Харків
Харківський національний
університет радіоелектроніки
2012

УДК 621.3

Сборник включен в список специальных изданий ВАК Украины по физико-математическим и техническим наукам.
Регистрационное свидетельство КВ № 12098-969 ПР от 14. 12. 2006.
Ответственность за содержание статей несут авторы.

Редакционная коллегия

В.М. Шокало, *д-р техн. наук, проф. (главный редактор)*
Ю.Б. Гимпилевич, *д-р физ.-мат. наук, проф.*
И.Д. Горбенко, *д-р техн. наук, проф.*
Ю.Е. Гордиенко, *д-р физ.-мат. наук, проф.*
А.Н. Довбня, *чл.-кор. НАНУ, д-р физ.-мат. наук., проф.*
В.М. Карташов, *д-р техн. наук, проф.*
А.А. Коноваленко, *академик НАНУ, д-р физ.-мат. наук*
В.М. Кузмичев, *д-р физ.-мат. наук, проф.*
Л.М. Литвиненко, *академик НАНУ, д-р физ.-мат. наук*
А.И. Лучанинов, *д-р физ.-мат. наук, проф. (зам. главного редактора)*
И.М. Неклюдов, *академик НАНУ, д-р физ.-мат. наук*
А.Г. Пащенко, *канд. физ.-мат. наук, доц. (ответственный секретарь)*
В.В. Поповский, *д-р техн. наук, проф.*
Э.Д. Прохоров, *д-р физ.-мат. наук, проф.*
А.И. Стрелков, *д-р техн. наук, проф.*
К.С. Сундучков, *д-р техн. наук, проф.*
П.Л. Токарский, *д-р физ.-мат. наук, проф.*
А.И. Фисун. *д-р физ.-мат. наук, проф.*
Г.И. Хлопов, *д-р техн. наук*
Я.С. Шифрин, *д-р техн. наук, проф.*

Международная редакционная коллегия

A.G. Karabanov, USA
S.E. Sandström, Sveden
N. Chichkov, Germany

Ответственный за выпуск И.Д. Горбенко, д-р техн. наук, проф.
Технический секретарь Е.С. Полякова

Рекомендовано Ученым советом Харьковского национального университета радиоэлектроники, протокол № 14 от 28.09.2012.

Адрес редакционной коллегии: Харьковский национальный университет радиоэлектроники (ХНУРЭ), просп. Ленина, 14, Харьков, 61166, тел. (0572) 7021-397.

*Сборник «Радиотехника» включен в Каталог подписных изданий Украины, подписной индекс **08391***

Перепечатка и использование материалов в любой форме возможны только с согласия редколлегии

© Харківський національний університет радіоелектроніки, 2012.

СОДЕРЖАНИЕ

БЛОЧНЫЕ СИММЕТРИЧНЫЕ ШИФРЫ

В.И. Долгов, А.А. Настенко О роли схем разворачивания ключей в атаках на итеративные шифры 7

В. И. Долгов, Е.Д. Мельничук S-блоки для современных шифров 16

А.В. Казимиров, Р.В. Олейников Криптоанализ шифра Mickey на основе анализа внутренних состояний 24

И.В. Лисицкая Вырожденные подстановки 31

В.И. Руженцев Про доказательство отсутствия эффективных байтовых дифференциальных характеристик для Rijndael-подобных шифров 39

А.А. Кузнецов, А.В. Коваленко, С.А. Исаев Формирование нелинейных узлов замен с использованием недвоичных криптографиических функций 43

КРИПТОГРАФИЧЕСКИЕ ПРЕОБРАЗОВАНИЯ

Е.Г. Качко, Д.С. Балагура, К.А. Погребняк, Ю.И. Горбенко Исследование методов вычисления инверсии в алгоритме NTRU 58

Ю.М. Лєншина, А.В. Лєншин Криптографічна підтримка послуг захисту від несанкціонованого доступу 64

К.А. Погребняк, Д.В. Повтарев Модель використання хмарних обчислень для задач• асиметричного криптоаналізу на прикладі факторизації чисел методом ρ-Полларда 72

І.Д. Горбенко, Л.В. Макутоніна Аналіз стійкості обчислювальних задач, що засновані на білінійних відображеннях 79

Д.В. Іваненко Методи протидії атакам спеціального виду на схеми направленого шифрування у кільцях зрізаних поліномів 90

И.Д. Горбенко, Р.И. Мордвинов Сущность и анализ криптографических требований стандарта NISTSP 800-90B 99

Е.В. Котух Универсальное хеширование с ограничением функционального поля алгебраических кривых 109

І. Д. Горбенко, М. В. Єсіна Алгоритм решета числового поля 116

В.Ю.Ковтун, А.А.Охрименко Подходы к распараллеливанию программной реализации операции умножения целых чисел 123

О.А. Шевчук, Ю.І. Горбенко Автентифікація апаратних засобів КЗІ 132

О.Г. Халимов, А.Н. Герцог Универсальное хеширование по обобщенным кривым Гурвица 140

А.П. Бубир, І.Д. Горбенко Оцінка стійкості направленого шифру NTRUдо атаки з адаптивно підібраними шифротекстами 147

Х.А. Бугаєнко, І.Д. Горбенко Аналіз двох методів автентифікації особи для застосування в електроному біометричному паспорті 152

ПЕРЕДАЧА И ОБРАБОТКА ИНФОРМАЦИИ

Ю.І. Горбенко, О.С. Тоцький, В.А. Пономар Аналіз методів знеособлення персональних даних 159

А.В. Потий, Д.Ю. Пилипенко Модель институционального управления деятельностью по защите информации 164

А.А. Замула, С.А.Сирота, Н.И. Косиковская Количественная оценка уязвимостей информационно-телекоммуникационных систем 171

А.А. Замула Предложения по построению широкополосных систем передачи со сложными сигналами 177

Б.О. Бабич, А.В. Сагун, О.А. Кожуховская, А.Д. Кожуховский Синтез системы многоуровневой защиты корпоративных порталов на платформе MS SHARE POINT 185

А.А. Смирнов Критерии и показатели эффективности стеганографических систем защиты информации 189

И.В. Олешко, И.Д. Горбенко Сравнительный анализ протоколов строгой аутентификации 198

РАДИОЛОКАЦИЯ

О.В. Сытник, В.М. Карташов, А.А. Супрун Пространственная селекция широкополосных источников по собственным числам ковариационной матрицы 210

О.О. Байздренко, В.Б. Лубський Розрахунок характеристик виявлення сигналів, відбитих об'єктами, розташованими на морській поверхні, полуактивною бістатичною РЛС з цифровим телевізійним сигналом освітлення 216

А.А. Буланый, Г.В. Майстренко, А.А. Стрельницкий, В.М. Шокало Сравнительный анализ помехозащищенности и спектральной эффективности Wi-Fi каналов связи с линейными и двумерными адаптивными антенными решетками при воздействии нескольких помех и одного сигнала 222

С.В Марченко, В.М Морозов., А.М. Съянов Исследование ФАР с диэлектрическим заполнениемем и согласующей периодической структурой 229

В.М. Карташов, Д.Н. Куля Синтез и анализ дискриминатора следящего устройства систем радиоакустического зондирования атмосферы 234

Б.В. Жуков, А.В. Одновол Контроль уровня легкоиспаряющихся жидкостей методом акустической локации 239

Г.М. Чекалин, Г.Н. Чекалина Об аналитических методах синтеза поляризационного эллипса 245

Ю.В. Лыков Результаты экспериментального исследования телефонных радиозакладных уст ройств методом нелинейной локации 252

В.И. Леонидов Модельно-структурный анализ эхосигналов акустического зондирования атмосферы 258

СИСТЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

И.С. Шостко, Ю.Э. Соседки, Алмакадма Таха Разработка рекомендаций по регулированию пропускной способности в WPAN 262

И.В. Ковтун Особенности передачи данных по блокам в транкинговой системе стандарта APCO 25 270

Ю.Г.Лега Завадостійкість m-позиційного автокореляційного приймача шумових сигналів в гауссовому каналі 275

С.А.Шейко Особенности межкадрового сжатия видеоинформации в устройствах видеонаблюдения и видеорегистрации 283

В.Г. Котух, М.А. Мирошник, С.Н. Селевко Методы планирования ресурсов в распределенных компьютерных системах 290

РАДИОТЕХНИЧЕСКИЕ УСТРОЙСТВА

Л.М. Карпуков, Р.Ю. Корольков Прямой метод синтеза полосно-пропускающих шлейфовых фильтров с чебышевской характеристикой 300

П.Ф.Лебедев, В.П.Дробышев Схемы замещения конденсаторов и катушек индуктивности 306

Д.Ю. Пенкин, Л.П. Яцук Анализ энергетических характеристик поперечной щели в широкой стенке прямоугольного волновода с локальным диэлектрическим включением 313

С. Л. Бердник Излучение электромагнитных волн электрически длинной щелью с диэлектрическим заполнением в узкой стенке многомодового прямоугольного волновода 322

О.І. Филипенко, О.В. Сичова Моделювання впливу структури фотонно-кристалічних волокон на розподіл модового поля та втрати оптичного сигналу в їх з'єднаннях 327

Н.И. Слипченко, В.А. Письменецкий, А.В. Фролов, Н.В. Герасименко, М.Ю. Гуртовой, Е.С. Глушко, Т.Е. Стыценко Исследование и оптимизация пленочных кремниевых аморфных фотопреобразователей на *p-i-n*-структурах 332

А.А. Жалило Разработка и тестирование новых эффективных методов и алгоритмов обнаружения и устранения фазовых скачков статических и кинематических ГНСС-наблюдений 340

ABSTRACTS 372

От редколлегии

В сборнике «Радиотехника» №170 были опубликованы три статьи П.П.Лезова. Редколлегия считает эти статьи ошибочными и приносит извинения за их публикацию.

Редколлегия

CONTENT

BLOCK AND SYMMETRIC CYPHERS

V.I. Dolgov, A.A. Nastenko On the role of key schedules in attacks on iterated ciphers 7
V.I. Dolgov, E.D. Melnichuk S-blocks for modern ciphers 16
O.V. Kazymyrov, R.V. Oliynykov Mickey cipher cryptanalysis based on internal states analysis 24
I.V. Lisitskaya Degenerated substitutions 31
V.I. Ruzhentsev About the proof of effective byte differential characteristics absence for Rijndael-like ciphers 39
A.A. Kuznetsov, A.V. Kovalenko, S.A. Isaev Synthesis of nonlinear substitution components with the use of non-binary cryptographic functions 43

CRYPTOGRAPHIC TRANSFORMATION

E.G. Kachko, D.S. Balagura, K.A. Pogrebnyak, Y.I. Gorbenko Investigation of methods for calculation of the inversion algorithm, NTRU 58
I.M. Lyenshyna, A.V. Lyenshyn Cryptographic support of unauthorized access protection services 64
K.A. Pogrebnyak, D.V. Povtariev Model of using cloud computing for problems of asymmetric cryptanalysis illustrated by the example of factorization of numbers by ρ-Pollard algorithm 72
I.D. Gorbenko, L.V.Makutonina Resistance analysis computational problems based on the bilinear maps 79
D.V.Ivanenko Methods to counteract side channel attacks to the schemes in rings of truncated polynomials 90
U.I. Gorbenko, R.I. Mordvinov Essence and analysis of cryptographic requirements from NIST SP 800-90B 99
Ye. Kotukh Universal hashing with limited functional field of algebraic curves 109
I. D. Gorbenko, M. V. Yesina Number field sieve algorithm 116
V.Y. Kovtun, A.O. Okhrimenko Approaches to parallelization of software implementation of integer multiplication 123
O.A.Shevchuk, U.I. Gorbenko HSM Authentication 132
O.G. Khalimov, A.N. Guetzog Universal hashing by generalized Hurwitz curves 140
A.P. Bubyr, I.D. Gorbenko Evaluation of lattice-based public key algorithm NTRU resistance against adaptive chosen cipher texts attack 147
A.P. Bubyr, I.D. Gorbenko Evaluation of lattice-based public key algorithm NTRU resistance against adaptive chosen ciphertexts attack 152

INFORMATION TRANSMISSION AND PROCESSING

Y.I. Gorbenko, A.S. Totsky, V.A. Ponomar Analysis of the methods of personal data anonymization 59
/ A.V. Potiy, D.Y. Pilipenko Institutional model of information security activities management 164
O.A.Zamula, S.O.Syrota, N.I.Kosikovskaia Quantitative assessment of information technology systems vulnerabilities 171
O.A. Zamula Proposal for construction of the wide-band transmission systems with complex signals 177
B.A. Babich, A.V. Sagun, O.A. Kozhukhovska, A.D. Kozhukhovskiy Synthesis of system of multilevel protection of corporate portals on the base of Share Point Platform 185
A.A. Smirnov Criteria and indexes of efficiency of the steganography systems of priv 189
/ I.V. Oleshko, I.D. Gorbenko Comparative analysis of strong authentication protocols 198

RADAR

O.V. Sytnik, V.M.Kartashov, O.O. Suprun Spatial selection of wide-band sources using eigenvalue of the covariance matrix 210
A.A. Baizdrenko, V.B. Lubsky Calculation of detection characteristics of signals, reflected by objects located on the sea surface, in semi-active bistatic radar with digital TV signal illumination 216
A.A. Bulany, G.V. Maistrenko, A.A. Strelnitsky, V.M. Shokalo Comparative analysis of immunity and spectral efficiency of Wi-Fi channels with linear and two-dimensional adaptive antenna array under the influence of several interferences and a signal 222

S.V. Marchenko, V.M Morozov, A.M Syanov Investigation of PAA with dielectrical inclusion and matching periodical structure 229
V.M.Kartashov, D.M.Kulia Synthesis and analysis of the tracker device discriminator of the atmosphere radio acoustic sensing systems 234
B.V. Zhukov, A.V.Odnovol Control over the level of easily evaporated liquids by the acoustic location method 239
G.M. Chekalin, G.N. Chekalina About analytical methods for synthesis of the polarization ellipse 245
Y.V. Lykov Results of experimental study of telephone radio bugs using nonlinear locations method 252
V.I. Leonidov Model-structural analysis of echo signals of atmospheric acoustic sounding 258

INFORMATION PROCESSING SYSTEMS

I.S. Shostka, J.E Sosedka, Almakadma Taha Development of recommendations concerning adjusting of carrying capacity in WPAN 262
I.V. Kovtun Particularities of data transmission to the blocks in the trunking system of APCO 25 standard 270
Yu.G. Lega Noise- immunity of m-positional autocorrelation receiver of noise signals in Gaussian channel 275
S. O. Sheiko Particularities of inter frame coding of video information in the video monitoring and video recording devices 283
V.G. Kotuh, M.A. Miroshnik, S.N. Selevko Methods of resources planning in the distributed computer systems 290

RADIO ENGINEERING DEVICES

L.M. Karpukov, R.Y. Korolkov The direct method of synthesis bandpass stub filters with Chebyshev characteristic 300
P.F. Lebedev, V.P. Drobyshev Equivalent circuits of condensers and inductance coils 306
D.Yu. Penkin, L.P. Yatsuk Analysis of power characteristics of the transversal slot in the wide wall of the rectangular waveguide with a local dielectric including 313
S. L. Berdnik Radiation of electromagnetic waves by electrically long slot with dielectric filling in the narrow wall of the multimode rectangular waveguide 322
A. I. Filipenko, O.V. Sychova Modeling of the photonic crystal fiber structure impact on the mode field distribution and on the optical signal loss in their connections 327
N.I. Slipchenko, V.A. Pismenetskiy, A.V. Frolov, N.V. Gerasimenko, M.Y. Gurtovoy, L.G. Glushko, T.E. Stytsenko Research and optimization of p-i-n structure based on the amorphous silicon films 332
A.A. Zhalilo Development and testing of new methods and algorithms for detecting and correction the carrier-phase slips of static and kinematic GNSS observations 340

ABSTRACTS 372

УДК 621. 3.06

В.И. ДОЛГОВ, д-р техн. наук; А.А. НАСТЕНКО

О РОЛИ СХЕМ РАЗВОРАЧИВАНИЯ КЛЮЧЕЙ В АТАКАХ НА ИТЕРАТИВНЫЕ ШИФРЫ

Ведение

Вопрос о роли схем разворачивания ключей в атаках на итеративные шифры давно считается одним их актуальных в криптографии. Название статьи повторяет название работы [1], подготовленной ещё в 2004 году, потому что хотим выразить свою точку зрения в этом вопросе, существенно отличающуюся от позиции авторов отмеченной работы, в которой рассматриваются итеративные шифры и их устойчивость к атакам линейного и дифференциального криптоанализа. Отмечается, что в теории эти нападения предполагают независимость от цикловых ключей шифров. Очень часто, однако, цикловые ключи вычисляются с помощью схем разворачивания ключей (ключевых расписаний) из короткого ключа в неслучайном режиме. С помощью экспериментов авторы показывают, что шифры со сложным ключевым графиком противостоят атакам линейного и дифференциального криптоанализа лучше, чем шифры с более простым ключевым графиком. Отмечается также известное в криптографической литературе положение, что с предположением независимых ключей вероятности дифференциалов и линейных оболочек могут быть смоделированы с помощью цепей Маркова и что для большинства таких шифров распределение вероятностей дифференциалов и линейных оболочек сходится к равномерному распределению после некоторого числа циклов. Представленные эксперименты с уменьшенными моделями шифра DES показывают, пишут авторы работы [1], что некоторые итеративные шифры с очень простым ключевым графиком никогда не достигают этого равномерного распределения. Кроме того, эксперименты показывают, что шифры с хорошо продуманным, сложным ключевым графиком достигают равномерного распределения быстрее (с использованием меньшего количества циклов), чем шифры с плохо разработанным ключевым графиком. В качестве побочного результата авторы отмечают, что существуют шифры, для которых дифференциал с наибольшей вероятностью для одного фиксированного ключа совпадает с дифференциалом с наибольшей вероятностью для любого другого ключа. Авторы считают, что это первый такой пример, приведенный в литературе. Имеются и другие публикации (например, [2, 3]), поддерживающие позицию авторов работы [1].

Цель данной статьи – ревизия изложенных выше результатов. Покажем, что как и в других наших работах [4, 5 и др.] показатели стойкости шифров к атакам линейного и дифференциального криптоанализа не зависят от схем разворачивания ключей. Кроме того, утверждение о том, что для любых ключевых графиков и большинства шифров, отнесенных к марковским, распределение вероятностей дифференциалов и линейных оболочек сходится к равномерному распределению после некоторого количества циклов, является неверным. Все шифры, как показано в [6, 7] являются марковскими и асимптотически приходят к показателям, свойственным случайным подстановкам. Наконец, покажем, что факт того, что существуют шифры, для которых дифференциал с наибольшей вероятностью для одного фиксированного ключа практически совпадает с дифференциалом с наибольшей вероятностью для любого другого ключа, является истиной для любого итеративного шифра. Более того, это свойство справедливо и для максимумов смещений линейных оболочек.

Эксперименты

Работа [1] строится на использовании уменьшенных версий шифров (8-битная, 10-битная, 12-битная). Но дело в том, что получить уменьшенную версию, повторяющую свойства большой, для шифра DES со слабым линейным преобразованием это нереализуемая

задача. Учесть нюансы, состоящие в том, что имеется слабость у третьего S-блока для дифференциального криптоанализа и слабость у 5-го S-блока для линейного криптоанализа в малой модели просто невозможно потому, что приходится иметь дело с шифром со значительно уменьшенным по размерам входным блоком данных, для которого переходы с особыми вероятностями, обнаруженными Э. Бихамом и А. Шамиром, а также М. Мацуи, как мы покажем, "тонут" в существенно более высоких значениях вероятностей переходов, свойственных малым моделям шифров.

Наконец, авторы скорее стремились подтвердить имеющиеся в литературе результаты и положения, а не проверить их объективность. Вот и получилось, что в этой работе представляются ожидаемые авторами результаты, которые, как мы покажем, оказались далёкими от реального состояния дел.

Мы в своих экспериментах сразу остановились на полной версии шифра DES. Повторили в экспериментах все пять вариантов схем разворачивания ключей, рассмотренных в работе [1], только переработали их для большой версии шифра.

Как и в [1] рассматривали поцикловые распределения максимумов полных дифференциалов, только строилась часть дифференциальной таблицы 64-битного шифра. Мы воспользовались нашей методикой анализа дифференциальных свойств больших моделей шифров в режиме их применения для зашифрования усечённых до 16-битных значений блоков данных, позволяющей построить эксперименты, находящиеся в рамках реализуемых вычислительных возможностей [8, 9].

Первая серия наших экспериментов выполнена с шифром DES в его оригинальном исполнении, т.е. была взята схема разворачивания ключей, полностью повторяющая её реализацию в самом стандарте.

В табл.1 представлены поцикловые значения максимумов полных дифференциалов (максимальных значений переходов таблиц XOR разностей) для шифра DES с различными вариантами использования "родной" схемы разворачивания ключей.

Таблица 1

Число циклов	"Родной" ключ	Независимые цикловые подключи	Слабый ключ	Нулевые подключи
1	65536	65536	65536	65536
2	65536	65536	65536	65536
3	211526	20736	8192	22528
4	388	400	334	5600
5	32	18	18	172
6	18	18	20	782
7	20	20	18	835
8	20	20	20	878
9	20	22	18	844
10	18	18	20	18
11	20	18	20	20
12	20	20	20	20
13	20	18	18	22
14	18	18	18	18
15	20	18	18	18
16	20	20	18	18

В первой колонке представлены значения дифференциалов при использовании в DES штатной схемы разворачивания ключей. Во второй колонке приведены данные для шифра, когда используется случайный набор цикловых подключей (схемы разворачивания ключей как таковой нет). В третьей колонке осуществляется зашифрование на слабом ключе: FEFEFE.... Видно, что в первых трёх случаях шифр DES после 5-ти 6-ти циклов зашифрова-

ния приходит к показателям случайной подстановки степени 2^{16} (становится случайной подстановкой). Слабый ключ (в шифре DES схема формирования цикловых подключей допускает ситуацию, когда во всех циклах алгоритма формируется один и тот же ключ) тоже не меняет общую ситуацию. Нулевые подключи в шифре DES затягивают переход к случайной подстановке до 10-ти циклов. Это просто означает, что алгоритм шифрования сам по себе не обладает хорошими перемешивающими свойствами.

В табл.2 представлены результаты второй серии экспериментов, где мы интересовались поцикловыми значениями максимумов смещений линейных оболочек шифра DES при различных вариантах использования схемы разворачивания ключей.

Таблица 2

Число циклов	"Родной" ключ	Независимые цикловые подключи	Слабый ключ	Нулевые подключи
1	32768	32768	32768	32768
2	32768	32768	32768	32768
3	20480	20480	20480	20480
4	2320	2952	3490	3206
5	810	845	840	783
6	870	803	807	821
7	805	815	859	791
8	860	839	824	802
9	823	809	848	802
10	779	839	816	835
11	803	773	828	828
12	802	789	817	854
13	782	799	829	831
14	835	814	811	804
15	878	876	852	799
16	844	832	811	816

В этом случае получается, что наличие или отсутствие ключей зашифрования не влияет на линейные свойства шифрующего преобразования. Во всех случаях шифры DES на пятом цикле зашифрования приходят к показателям случайной подстановки степени 2^{16} (становятся случайными подстановками).

Уже из этих результатов становится понятно, что в рамках наших экспериментов различные схемы разворачивания ключей не повлияют на дифференциальные и линейные свойства шифра. Но мы добросовестно повторили все последующие эксперименты работы [1].

Как и в работе [1], были использованы пять вариантов построения схем разворачивания ключей. Кратко напомним эти решения с учётом их модификации применительно к большому шифру DES.

В качестве входных данных для четырех названных основными ключевых графиков теперь уже рассматривается 96-битный ключ K, из которого формируется r 48-битных битных цикловых подключей K_i для $i = 1, ..., r$. Все четыре алгоритма берут выбранный пользовательский ключ и разделяют его на две 48-битные половины K^L и K^R.

Первый ключевой график строится следующим образом:

Key schedule 1:
Input: $K = K^L \mid K^R$
For i =1 **to** r/2 **do**
$K_{2i-1} = K^L$
$K_{2i} = K^R$
For $i = 0$ **to** r **do** $K_i = K_i$ XOR i.

Здесь цикловые подключи конструируются простым повторением отобранных пользователем ключевых половинок для каждого из циклов.

Хорошо известно, отмечают авторы работы [1], что такие ключевые графики оставляют шифр очень уязвимым для так называемых связанных ключевых атак [10, 11], а также слайд атак [12]. Тем не менее, это ключевое расписание, считают они, является слабым и в том смысле, что циклы шифра с четными номерами зависят только от одной ключевой половины, а циклы шифра с нечетными номерами зависят только от второй ключевой половины.

Чтобы избежать этой симметрии второй ключевой график использует ключевые половины в циклах в ином порядке:

Key schedule 2:
Input: $K = K^L \mid K^R$
For i =1 **to** r/4 **do**
$K_{4i-3} = K^L$
$K_{4i-2} = K^R$
$K_{4i-1} = K^R$
$K_{4i} = K^L$
For i =0 **to** r **do** $K_i = K_i$ XOR i

Как следует из алгоритма, к цикловым ключам добавляются цикловые константы. Первые два графика используют 48-битные половинки K непосредственно, то есть наименее значимый бит циклового ключа зависит только от наименее значимого бита двух половинок входного ключа.

Чтобы избежать таких свойств, третий график использует сдвиги (ротации) чтобы распространять влияние битов ключа K по всем позициям цикловых подключей:

Key schedule 3:
Input: $K = K^L \mid K^R$
$K_1 = K^L$
$K_2 = K^R$
K_3 = LeftShift(K^L, 24) + RightShift(K^R, 24)
K_4 = LeftShift(K^R, 24) + RightShift(K^L, 24)
For i = 5 **to** r **do** K_i = Rotate(K_{i-3},12)
For i =1 **to** r **do** $K_i = K_i$ XOR i

Leftshift принимает 24 младших бита на вход и сдвигает их на 24 позиции влево. *Right-Shift* принимает 24 старших бита на вход и сдвигает их на 24 позиции вправо. Как следствие, третий цикловой подключ K_3 зависит от 24 битов из K^L и 24 битов из K^R, в то время как четвертый K_4 цикловой подключ зависит от остальных 24 битов K^L и K^R. Оставшиеся цикловые ключи генерируются как сдвинутые версии предыдущих цикловых ключей. Чтобы избежать тривиальной симметрии и слабых ключей, используются упорядоченные цикловые константы (различные для всех цикловых подключей).

Четвертый график еще сложнее. Здесь ряд (серия) цикловых подключей TK_1, ..., TK_r генерируются аналогично предыдущему. Затем эти цикловые ключи используются в шифре, о котором идёт речь, чтобы создать (реальные) цикловые ключи для экспериментов. Шифр используется в режиме счетчика и результирующие шифртектовые половины являются исключительно упорядоченными, для генерации (реальных) цикловых подключей .K_1,..., K_r:

Key schedule 4:
Input: $K = K^L \mid K^R$
$TK_1 = K^L$
$TK_2 = K^R$
$TK_3 = K^L$ XOR K^R
For i = 4 **to** r **do** TK_i = Rotate(TK_{i-3},1)

For i =0 **to** r **do** $TK_i = TK_i$ XOR i
TK:= { TK_1,..., TK_r}
For i =1 **to** r **do**
$C = (C^L \mid C^R)$ = encrypt(i,T_K)
$K_i = C^L$ XOR C^R

Пятый ключевой график просто использует независимые цикловые ключи, то есть для тестирования шифра выбранный пользователем ключ в общей сложности имеет $48r$ бит. Это "ключевой график" уже рассмотрен в первой серии экспериментов.

Результаты оценки влияния четырех представленных выше основных ключевых графиков на дифференциальные и линейные показатели шифра DES иллюстрируют табл. 3 и 4.

Табл. 3 иллюстрирует результаты третьей серии экспериментов при использовании в шифре DES приведенных выше четырёх схем разворачивания ключей. В этой серии экспериментов изучались дифференциальные показатели схем зашифрования.

Таблица 3

Число циклов	Схема разворачивания 1	Схема разворачивания 2	Схема разворачивания 3	Схема разворачивания 4
1	65536	65536	65536	65536
2	36864	43008	43008	43008
3	12908	12350	25392	21128
4	302	132	114	118
5	18	18	20	18
6	20	18	20	18
7	20	20	20	18
8	18	18	20	20
9	20	18	18	18
10	18	20	18	18
11	20	20	18	20
12	18	18	20	20
13	20	20	20	20
14	20	20	18	18
15	20	20	18	20
16	20	20	20	20

Для всех перечисленных ключевых графиков перебор был реализован так, чтобы найти все дифференциалы и линейные оболочки для различного числа циклов. Для r-цикловой версии шифра и для каждого ключевого графика эксперименты проводились следующим образом:

Для каждого значения ключа были вычислены все r-цикловые дифференциалы и все r-цикловые линейные оболочки (корпуса). Корпус и дифференциал с наибольшей вероятностью по всем входам записывались. В наших экспериментах в отличие от работы [1] все расчёты выполнялись на одном из вариантов ключа зашифрования. Здесь уже учитывался опыт выполнения подобных исследований [4, 5], в соответствии с которым результаты, как для максимумов дифференциалов, так и для максимумов смещений практически не зависят от ключа зашифрования. Среднеквадратическое отклонение не превышает значения ±1,5 для максимумов дифференциалов и ± 35 для максимумов смещений.

Табл. 4 иллюстрирует результаты четвёртой серии экспериментов при использовании в шифре DES приведенных выше четырёх схем разворачивания ключей. В этой серии экспериментов изучались линейные показатели схем зашифрования.

Во всех рассмотренных случаях получается, что различные схемы разворачивания ключа на уровне оценки дифференциальных и линейных показателей 16-битных переходов для

шифра DES приводят к одним и тем же результатам. Через пять-шесть циклов все варианты шифров приходят к показателям случайной подстановки степени 2^{16}.

С другой стороны, известно, что большой шифр DES приходит к показателям случайной подстановки не через пять циклов, а гораздо позже. (Вспомним дифференциальную 13-цикловую характеристику для входной разности "1960", на которую построили атаку Э. Бихам и А. Шамир.)

Просто в 16-битном варианте построения дифференциальной характеристики, использованном в работе, выделить этот особый переход не представляется возможным. Максимум перехода при 13 циклах, который соответствует вероятности $2^{-47,2}$, равен значению $2^{16,8} = 114104,8$ (это для дифференциальной таблицы размером $2^{64}\times2^{64}$). Для фрагмента таблицы $2^{32}\times2^{32}$ ожидаемое значение максимума будет равно $\sqrt{114105} \approx 337,8$. а для фрагмента таблицы размером $2^{16}\times2^{16}$ значение максимума соответственно получается равным $\sqrt{338} \approx 18$, что на уровне ожидаемого значения максимума случайной подстановки степени 2^{16} для особенного перехода "1960" уже не заметишь.

Таблица 4

Число циклов	Схема разворачивания 1	Схема разворачивания 2	Схема разворачивания 3	Схема разворачивания 4
1	32768	32768	32768	32768
2	32768	32768	32768	32768
3	22528	14336	17664	20480
4	2744	4176	3464	2508
5	806	797	864	824
6	776	810	824	829
7	818	811	825	835
8	805	787	811	831
9	844	845	818	809
10	794	799	860	843
11	825	816	857	886
12	828	817	815	808
13	871	837	817	788
14	812	804	847	838
15	799	847	810	846
16	834	834	894	824

В этом видится один из недостатков использования малых моделей, который проявляется в нивелировании переходов, превышающих стандартные значения максимумов дифференциальной таблицы большого шифра множеством значений переходов, приходящихся на существенно уменьшенную её часть. В данном случае из всей таблицы размером $2^{64}\times2^{64}$ рассматривается только часть размером $2^{16}\times2^{16}$.

Ещё одно объяснение этого эффекта. Можно реализовать малую (16-битную) модель шифра DES с первыми тремя "родными" S-блоками, повторяющую особенности реализации большого шифра DES. Цикловая функция в этом случае строится так. Правая половинка из восьми бит входного блока данных дополняется расширением E до 12 битов, так что на входы трёх S-блоков цикловой функции поступают биты в таком порядке 12, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11. На вход первого S-блока поступают 12, 1, 2, 3, 4, и 5-й биты, на вход второго S-блока поступают 4, 5, 6, 7, 8, и 9-й биты и на вход третьего S-блока поступают 8, 9, 10, 11, 12, и 1-й биты. В этом наборе из 12 битов повторяющимися являются 4 и 5-й биты, 8 и 9-й биты, 12 и 1-й биты (всего получается восемь различных битов входа). Далее идут, как и положено, S-блоки и P подстановка. С такой цикловой функцией можно реализовать переход 1960 и для малого шифра с вероятностью 1/234, свойственный "большому" DES.

Тогда для 16-битного шифра вероятность перехода 1/234 за первые два цикла приводит к значению перехода на выходе равному 280; вероятность 1/234 для первой и второй пары циклов даёт уже значение перехода 1,19, что значительно меньше двадцатки. В результате. получается, что и для малой модели уже после четырех циклов значение особого перехода "тонет" в значениях других переходов. Малая модель, как и рассмотренная выше большая для 16-битных переходов становится случайной подстановкой после 5-6-ти циклов независимо от схемы разворачивания ключей.

Тем более нельзя зафиксировать отмеченную выше особенность реализации шифра DES в уменьшенных до 8-10 битовых входов моделях, рассмотренных в работе [1]. Таким образом, в наших исследованиях мы полностью повторили эксперименты работы [1].

Вспомним теперь о большом шифре DES.

Мы нашли ещё две работы [14, 15], связанные с изучением возможностей реализации атак дифференциального криптоанализа на шифр DES. В работе [14] отмечено, что вероятность 1/234 является только средней вероятностью. Более того, наряду с входной разностью $\Phi = 19600000_x$ (в наших обозначениях это разность "1960") существует и входная разность $\Gamma = 1b600000_x$ с такой же средней вероятностью перехода. Эту возможность описывали Э. Бихам и А. Шамир в работе [16]. В этой работе, на которую есть ссылка и в работе [14], указано, что если шестой ключевой бит, используемый в S2, отличается от второго ключевого бита, используемого в S3, то вероятность для входа Φ увеличивается до 1/146, в то время как вероятность для входа Γ уменьшается до 1/585. Если же эти два ключевых бита равны, то результирующие вероятности меняются местами.

Авторы работы [14] назвали эти ключевые биты *критическими* для Φ и Γ. В своей атаке на DES [17] Бихам и Шамир использовали эти две характеристики, чтобы построить 13-цикловые характеристики, где шесть циклов имеют входной xor Φ или Γ. Вероятность, как утверждается, должна быть $(1/234)^6 \approx 2^{-47,22}$. Но в зависимости от значений шести пар критических ключевых битов вероятность для Φ будет варьироваться в пределах от $(1/146)^6 \approx 2^{-43,16}$ до $(1/585)^6 \approx 2^{-55,16}$ и, наоборот, для Γ.

И тут дело не в специально подобранной схеме разворачивания ключей. Э. Бихам и А. Шамир в своей работе полагали ключевые биты, воздействующие на S-блоки статистически независимы. Их методика вычисления вероятности двухцикловой характеристики обнуляющего типа основывалась на допущении, что после сложения расширенного входного полублока данных с цикловым подключом (биты которого в пределах подключа независимы) входы двух смежных (соседних) S-блоков также становятся "статически" независимыми [15]. Однако более тщательный анализ F функции показывает (это мы цитируем работу [15]), что использование в шифре DES сложения с подключом после E расширения приводит к тому, что условия "сшивки" S-блоков становятся зависимыми от ключевых битов, о чём говорится в работе [14]. Получается, что даже в оригинальной разработке DES при использовании "родной" схемы разворачивания ключей возможны семь вариантов вероятностей в указанном выше диапазоне, а каждое из предельных значений вероятностей возможно на 2^{50} ключей, т.е. на каждом 64-м ключе. Значение 2^{-47}, использованное Э. Бихамом, может быть получено как среднее от двух предельных значений:

$$(1/146)\cdot 1/2 + (1/585)\cdot 1/2 = 1/234.$$

При организации атаки нападающий не знает значения ключа зашифрования и вынужден ориентироваться на среднее значение характеристики, а это значение не зависит от значений цикловых подключей. Представляется, что этот же эффект буде наблюдаться и при иных схемах разворачивания ключей. Тем не менее, можно заключить, что для шифра DES его стойкость к атакам, по крайней мере дифференциального криптоанализа, зависит от схемы разворачивания ключей (от ключевого графика).

При любых вариантах схем разворачивания ключей шифр рано или поздно приходит к показателям случайной подстановки.

Так, если исходить из того, что для 13-ти циклов мы имеем наибольшую вероятность максимального перехода $2^{-47,22}$, то ещё одна дополнительная двухцикловая характеристика (15 циклов) приводит к значению максимального ожидаемого перехода 114104,8/234 = = 487,63, а для 17 циклов имеем значение перехода 487,63/234 = 2,08 – значение существенно меньшее максимума полного дифференциала случайной подстановки степени 2^{64} (это значение максимума близко к 68 [18]. На 16-ти циклах шифрования шифр DES становится случайной подстановкой (здесь ведём вычисления с дифференциальной характеристикой, а не с полным дифференциалом, который по значению перехода должен превосходить дифференциальную характеристику). Даже при использовании двухцикловой характеристики с вероятностью 1/146 (специально подобранные цикловые подключи) шифр DES на 16-ти циклах приобретает ожидаемое значение максимума дифференциального перехода равное 88, что всё равно очень близко к 68.

Таким образом, можно сделать вывод о том, что схемы разворачивания ключей (исключая специально подобранные варианты) для полного шифра DES никакой криптографической значимости (по отношению к показателям стойкости к атакам дифференциального криптоанализа) не имеют. Независимо от того, присутствует ли какая-либо схема разворачивания ключей в шифре или цикловые ключи имеют нулевые значения, шифр DES после шестнадцати циклов приходит к показателям случайной подстановки. Особые переходы, имеющиеся в шифре DES, на наш взгляд, не имеют криптографической ценности, так как являются принадлежностью только этого шифра. Атака на шифр DES строится на использовании укороченная дифференциальная характеристика с переходами обнуляющего типа (специфическими именно для шифра DES), когда шифр ещё не достиг установившегося (стационарного) значения максимума полного дифференциала.

Одновременно полученные результаты подтверждают практическую независимость установившихся законов распределения полных дифференциалов шифра и распределения смещений таблицы линейных аппроксимаций от ключей шифрования.

Если вспомнить теорему [13] о том, что с момента прихода шифра к стационарному состоянию дальнейшее наращивание числа циклов не меняет закона распределения полных дифференциалов (смещений ЛАТ) на его выходе, то становится понятным, что результирующее распределение полных дифференциалов, как и результирующее распределение смещений линейных оболочек, никогда не приходит к значению $p = 2^{-64}$.

Заметим далее, что имеющаяся в шифре DES особенность, позволяющая реализовать эффект накопления вероятностей на отдельных переходах, связана с возможностью реализации в шифре характеристик обнуляющего типа, когда ненулевая разность на входе преобразования переходит в нулевую разность на выходе (когда удаётся при формировании дифференциальных характеристик использовать переходы $0 \rightarrow 0$ без потери вероятностей). Условиями возникновения такой ситуации являются:

1) использование в шифре несимметричных S-блоков, допускающих переходы входной ненулевой разности в нулевую выходную;

2) применение петлевой схемы формирования цикловых переходов, при которой удаётся реализовать попеременное чередование активных и пассивных циклов;

Очевидно, что эти условия не выполняются для шифров с квадратными S-блоками (для SPN шифров и ряда известных Фейстель подобных шифров: Камелия, ГОСТ 28147-89 и др.). Шифр DES выступает здесь как заметное исключение.

Выводы

По результатам экспериментов все схемы разворачивания ключей для шифра DES приводят к одному и тому же результату. Это значит, что, по крайней мере, на уровне 16-битных дифференциальных показателей схемы разворачивания ключей для шифра DES не оказывают существенного влияния на стойкость к атакам дифференциального и линейного криптоа-

нализа. Имеющиеся особые переходы в шифре DES в той или иной мере будут характерны для схем разворачивания ключей всех конструкций.

Показатели стойкости шифра зависят не от схем разворачивания ключей, а от значений используемых ключей, от присутствия в цикловых подключах критических битов. Среднее значение вероятностей дифференциальных характеристик, по-видимому, будет одним и тем же для разных схем разворачивания ключей.

Все итеративные шифры, как следует из развиваемой в [13] методологии оценки стойкости в БСШ к атакам дифференциального и линейного криптоанализа, после небольшого начального числа циклов зашифрования (в наших экспериментах для DES после 5-6, а для полного DES после 16 циклов) приходят к стационарным состояниям, свойственным случайным подстановкам соответствующих степеней, которые сохраняются при наращивании числа циклов, и, следовательно, предельные распределения не являются равномерными.

Список литературы: 1. *Lars R. Knudsen1 and John E. Mathiassen.* On the Role of Key Schedules in Attacks on Iterated Ciphers. P. Samarati et al. (Eds.): ESORICS 2004, LNCS 3193, pp. 322–334, 2004. Springer-Verlag Berlin Heidelberg 2004. 2. *Uri Blumenthal and Steven M. Bellovin,* A better Key Schedule for DES-like ciphers. Proceedings of Pragocrypt'96 30 September – 3 October 1996. 3. *Лепеха, А.Н.* Сравнительный анализ схем разворачивания ключей блочных симметричных шифров / А.Н. Лепеха // Радиотехника. – 2005. – Вып. 141. – С. 64 -69. 4. *Долгов, В.И.* Дифференциальные свойства блочных симметричных шифров, представленных на украинский конкурс / В.И. Долгов, А.А. Кузнецов, С.А. Исаев // Электронное моделирование. – 2011.– Т 33, № 6. – С. 81-99. 5. *Кузнецов, А.А.* Линейные свойства блочных симметричных шифров, представленных на украинский конкурс / А.А. Кузнецов, И.В. Лисицкая, С.А. Исаев // Прикладная радиоэлектроника. – 2011. – Т.10. №2 – С. 135-140. 6. *Лисицкая, И.В.* блочные симметричные шифры и марковские процессы / И.В. Лисицкая // Прикладная радиоэлектроника. – 2012. – Т. 11, № 2 – С. 137-143. 7. *Лисицкая, И.В.* Оценки максимальных значений дифференциалов и линейных корпусов Марковских шифров / И.В. Лисицкая., В.И. Долгов, А.А. Настенко // Прикладная радиоэлектроника. – 2012. – Т. 11, № 2 – С. 144-151. 8. *Лисицкая И.В.* Большие шифры – случайные подстановки / И.В. Лисицкая, А.А. Настенко // Радиотехника. – 2011. – Вып. 166. – С. 50-55. 9. *Лисицкая, И.В.* Дифференциальные свойства шифра FOX / И.В. Лисицкая, Д. С. Кайдалов // Прикладная радиоэлектроника. – 2011. – Т. 10, № 2. – С. 122-126. 10. *Biham, E.* New types of cryptanalytic attacks using related keys. In T. Helleseth, editor, Advances in Cryptology: EUROCRYPT'93, Lecture Notes in Computer Science 765, pages 398–409. Springer Verlag, 1993. 11. *Knudsen, L.R.* Cryptanalysis of LOKI'91. In J. Seberry and Y. Zheng, editors, Advances in Cryptology, AusCrypt 92, Lecture Notes in Computer Science 718, pages 196–208. Springer Verlag, 1993. 12. *Biryukov, A. and Wagner, D.* Slide attacks. In L. R. Knudsen, editor, Fast Software Encryption, Sixth International Workshop, Rome, Italy, March 1999, Lecture Notes in Computer Science 1636, pages 245–259. Springer Verlag, 1999. 13. *Лисицкая,И.В.* Методология оценки стойкости блочных симметричных шифров / И.В. Лисицкая // Автоматизированные системы управления и приборы автоматики. – 2011. – № 163. – С. 123-133. 14. *Knudsen, L.R.* Iterative characteristics of DES and s^2-DES, Advances in Cryptology, Proc. Crypto '92, LNCS 740, E. F. Brickell, Ed., Springer-Verlag, 1993, pp. 497-511. 15. *Головашич, С.А.* Ключевые группы в атаках дифференциального криптоанализа DES-подобных шифров / С.А. Головашич // Радиотехника.– 2000. – Вып. 114. – С. 57-62. 16. *Eli Biham, Adi Shamir.* Differential Cryptanalys is of DES- like Cryptosystems. Journal of Cryptology, Vol. 4 No. 1 1991. 17. *Eli Biham, Adi Shamir.* Differential Cryptanalysis of the full 16- round DES. Technical Report # 708, Technion – Israel Institute of Technology. 18. *Олейников, Р.В.* Дифференциальные свойства подстановок / Р.В. Олейников, О.И. Олешко, К.Е. Лисицкий, А.Д. Тевяшев // Прикладная радиоэлектроника. – 2010. – Т.9. – № 3. – С. 326-333.

Харьковский национальный университет радиоэлектроники

Поступила в редколлегию 25.09.2012

УДК 621. 3.06

В.И. ДОЛГОВ, д-р техн. наук, Е.В. МЕЛЬНИЧУК

S-БЛОКИ ДЛЯ СОВРЕМЕННЫХ ШИФРОВ

Введение

В предыдущей работе [1] было показано, что S-блоки для блочных симметричных шифров с хорошими криптографическими показателями можно выбрать на основе случайного отбора. Для байтовых S-блоков это сделать совершенно просто, так как с высокой вероятностью случайно выбранный S-блок будет обладать криптографическими свойствами, не уступающими лучшим известным конструкциям S-блоков. Здесь имеется в виду новый показатель оценки эффективности S-блоковых конструкций введенный в работе [1], который строится на основе оценки числа циклов шифра, необходимых ему для прихода к стационарному состоянию, свойственному случайной подстановке. Более совершенными будут считаться те S-блоки, с которыми шифр приходит к стационарным значениям максимумов полных дифференциалов и максимумов смещений линейных оболочек, характерных для случайной подстановки, за меньшее число циклов

В данной статье ставится задача подтвердить это положение на примере отбора подстановок (случайных S-блоков) для известного мирового лидера – шифра Rijndael и шифров, представленных на прошедший украинский конкурс по выбору претендента на национальный стандарт шифрования.

Методика выполнения исследований

Методика определения дифференциальных и линейных показателей полноразмерных шифров подробно представлена в работах [2, 3 и др.]. Основой этой методики является определение поцикловых значений максимумов 16-битных разностей (полных дифференциалов) и 16-битных значений максимумов смещений линейных оболочек для сегментов блоков данных на входах и выходах шифров с разными конструкциями S-блоков с последующим сравнением между собой этих решений по числу циклов, необходимых для прихода каждого из шифров к показателям случайной подстановки, т.е. практически строятся существенно уменьшенные части таблиц полных дифференциалов и смещений таблиц линейных аппроксимаций больших шифров, и на основе сравнительной оценки показателей этих усеченных таблиц принимается решение о преимуществах того или иного варианта построения шифра, а значит, и S-блоков, использованных в нем.

S-блоки для шифра Rijndael

Далее излагаются результаты вычислительных экспериментов. В табл. 1 представляются поцикловые значения максимумов полных дифференциалов для шифра Rijndael с различными S-блоками. В первой колонке приведены данные для шифра Rijndael с использованием стандартных S-блоков. В колонках под номерами 2 – 6 в шифре Rijndael используются случайно порожденные S-блоки, причем при отборе случайных S-блоков на них не накладывались никакие ограничения. Взяты подряд пять случайно сгенерированных S-блоков.

Как было отмечено, в экспериментах строились дифференциальные таблицы для 16-битных разностей на входе и выходе шифров. Использовались разности для 4-го и 7-го байтов входа и выхода.

В качестве мастер-ключей взяты случайно порожденные 128-битные блоки, приведенные в табл. 2.

Во всех случаях использовалась стандартная схема разворачивания ключа. Общий вывод из представленных результатов состоит в том, что подряд взятые случайные S-блоки обеспечивают для шифров дифференциальные показатели, практически не уступающие показателям шифра с "родными" S-блоками, отобранными по специальным условиям. Полу-

чается, что "охота" за S-блоками с улучшенными криптографическими показателями, ведущаяся активно в криптографической литературе, не имеет смысла.

Таблица 1

Число циклов	Подстановки					
	1	2	3	4	5	6
1	65536	65536	65536	65536	65536	65536
2	24	18	20	18	20	20
3	20	18	18	20	18	20
4	18	18	18	18	20	18
5	18	18	20	18	18	20
6	20	20	18	22	18	20
7	18	20	20	20	18	20
8	20	18	20	20	20	20
9	20	20	20	18	18	20
10	18	18	20	20	20	18
11	18	18	18	18	18	18
12	20	20	20	20	20	20
13	18	18	20	20	18	20

Таблица 2

№	Мастер-ключи
1	B7CC40EC614A6410965517B40CDEEB5 ADD76369264B5AC3EA16D2FF8DDC1 5
2	A97F2F19A1115A3833E6C7D3126E8D5 A1FA8BCC3A5E963A13818B9E1DFFE4 5A
3	5F21A0D64CDEC949A41FB7B1EB8C66 E2C07BC7F8314C3BBC22D2D403924C2 A5
4	C0F3765F31666EF6F86F7B1AD9DECA8 926AF75B438A9A8D4C1FB71822CAE1 F
5	91E561CDAB2E2A88D53360BDB3BF26 5333EEB580DCAFCFAC31173488A1557 C
6	EF25ACE0553F879F4DF48E14B6AD650 A6308253B850C9B5526FC2A506AD

Для более полного отражения криптографических свойств шифров в табл. 3 и 4 представлены поцикловые законы распределения переходов таблиц полных дифференциалов для шифра Rijndael с родной подстановкой (табл. 3) и этого же шифра – со случайной подстановкой 1 (табл.4). Из табл. 3 и 4 хорошо видно, что законы распределения вероятностей, начиная уже с первого цикла, очень близки друг к другу.

В табл. 5 представлены конкретные показатели подстановок S-блоков (одинаковых) шифра Rijndael.

В табл. 6 для сравнения приведены характеристики случайного S-блока под номером 2 в табл. 1 (этот же S-блок приведен под номером 1 в табл.2). Сравнение результатов табл. 5 и 6 показывает, что случайный S-блок по комбинаторным показателям оказывается близким к специально отобранному S-блоку шифра Rijndael, в то время как законы распределения переходов и смещений соответствующих таблиц заметно отличаются друг от друга.

БИБЛИОТЕКА ХНУРЭ Инв. № 827248

Далее приведены результаты экспериментов с шифром Rijndael, но теперь для него строятся поцикловые распределения максимумов линейных оболочек.

В табл. 7 в первой колонке приведены показатели шифра для стандартной подстановки (табл. 5), а в колонке 2 используется случайно сгенерированный S-блок из табл. 6.

И в этом случае результаты свидетельствуют о том, что шифр Rijndael со случайными S-блоками практически не уступает по эффективности (по числу циклов, необходимых для перехода к показателям случайной подстановки) оригинальной разработке. Для получения 16-битного входа/выхода были взяты 4-й и 7-й байты входного и выходного блоков полноразмерного шифра AES с размером блока 256 бит и длиной ключа 256 бит.

Таблица 3

Число циклов	Rijndael с родной подстановкой. Числа ячеек таблицы полных дифференциалов по их заполнениям												
	0	2	4	6	8	10	12	14	16	18	20	22	512
1	4286545665	0	0	0	0	0	0	0	0	0	0	0	$8{,}3\cdot10^{6}$
2	2604966614	$1{,}30\cdot10^{9}$	325651069	54255750	6775997	679213	56358	4109	262	3	1	0	0
3	2604978428	$1{,}30\cdot10^{9}$	325613734	54269116	6785299	678244	56157	3982	265	21	1	0	0
4	2605010786	$1{,}30\cdot10^{9}$	325648583	54268197	6783670	678427	56728	4125	243	15	0	0	0
5	2604983918	$1{,}30\cdot10^{9}$	325605055	54275819	6784332	678419	56753	4021	234	18	0	0	0
6	2605008408	$1{,}30\cdot10^{9}$	325617103	54280422	6787319	677529	56344	3932	287	20	3	0	0
7	2605000214	$1{,}30\cdot10^{9}$	325623218	54272761	6784787	679016	56636	4159	269	19	0	0	0
8	2604947277	$1{,}30\cdot10^{9}$	325616851	54258309	6782134	677200	56351	4042	226	16	1	0	0
9	2604952924	$1{,}30\cdot10^{9}$	325621156	54255725	6783272	677706	56566	3973	287	14	1	0	0
10	2604968362	$1{,}30\cdot10^{9}$	325629098	54261284	6781522	678581	56346	3926	224	9	0	0	0
11	2604965638	$1{,}30\cdot10^{9}$	325627806	54261335	6783337	676394	56550	4012	247	18	0	0	0
12	2604962347	$1{,}30\cdot10^{9}$	325627417	54258891	6780754	678731	56652	4006	240	11	2	0	0
13	2604987639	$1{,}30\cdot10^{9}$	325617247	54268082	6785738	679260	56582	4035	266	14	0	0	0

Таблица 4

Число циклов	Rijndael с родной подстановкой, 4-й и 7-й байт входа/выхода. Числа ячеек таблицы полных дифференциалов по их заполнениям														
	0	2	4	6	8	10	12	14	16	18	20	22	512	1024	65536
1	4286552577	0	0	0	0	0	0	0	0	0	0	0	$8\cdot10^{6}$	6912	255
2	2605185259	302205809	325640895	54324835	6802164	681450	56881	4169	282	15	1	0	0	0	0
3	2604968650	302524923	325615136	54279227	6776220	676897	56586	3862	243	15	1	0	0	0	0
4	2604914122	302616478	325600024	54254205	6778669	677154	56752	4077	269	10		0	0	0	0
5	2604976365	302521578	325612992	54265532	6785419	678839	56688	4073	256	16	2	0	0	0	0
6	2604993128	302492271	325620412	54270903	6785308	679123	56355	3999	244	16	1	0	0	0	0
7	2604997373	302481170	325628952	54269811	6785059	678762	56371	4009	240	12	1	0	0	0	0
8	2604965272	302523997	325631025	54264889	6778260	677406	56600	4010	285	14	2	0	0	0	0
9	2604998699	302493460	325603183	54279388	6787784	678264	56710	4018	239	14	1	0	0	0	0
10	2604983934	302504978	325619403	54272514	6780828	679031	56728	4047	283	14		0	0	0	0
11	2604978318	302507735	325630814	54263127	6782379	678333	56715	4056	274	9		0	0	0	0
12	2604972063	302524389	325616253	54266366	6783979	677808	56651	3997	243	8	3	0	0	0	0
13	2604965308	302531402	325621514	54262662	6781353	678683	56526	4028	272	11	1	0	0	0	0

Таблица 5

Количество циклов: 5 Количество инверсий: 16753 Количество возрастаний: 126
Максимум таблицы XOR: 4 Количество максимумов таблицы XOR: 255 Распределение элементов таблицы XOR 0: 32640 2: 32130 4: 255
Максимум таблицы LAT: 16 Количество максимумов таблицы LAT: 1275 Распределение элементов таблицы LAT 0: 4080 2: 12240 4: 9180 6: 10200 8: 8670 10: 6120 12: 9180 14: 4080 16: 1275

Таблица 6

Количество циклов: 5 Количество инверсий: 16849 Количество возрастаний: 130	
Максимум таблицы XOR: 10 Количество максимумов таблицы XOR: 9 Распределение элементов таблицы XOR	
0: 39277 2: 19865 4: 4998	6: 779 8: 100 10: 9
Максимум таблицы LAT: 34 Количество максимумов таблицы LAT: 1 Распределение элементов таблицы LAT	
0: 6633 2: 12046 4: 11711 6: 9640 8: 8119 10: 5872 12: 4276 14: 2784 16: 1799	18: 1035 20: 575 22: 281 24: 135 26:75 28:30 30: 10 32: 3 34: 1

Таблица 7

Число циклов	Стандартная подстановка	Случайная подстановка
1	11264	10752
2	852	804
3	807	812
4	805	812
5	843	796
6	874	800
7	808	806
8	876	824
9	810	815
10	817	840
11	840	827
12	811	843
13	818	811

Из табл. 7 видно, что законы распределения вероятностей, начиная с первого цикла, очень близки друг к другу.

Рассмотрим теперь, как будут вести себя со случайными S-блоками шифры украинского конкурса. Здесь будем рассматривать три шифра: Калину, Мухомор и Лабиринт [4 – 6]. У шифра ADE в процессе экспертизы были обнаружены слабости [7], поэтому здесь исключим его из рассмотрения.

S-блоки для шифра Калина

В шифре Калина [5] используются восемь разных S-блоков, отобранных по специальным требованиям, о которых в спецификации не сообщается. Результаты тестирования шифра Калина с S-блоками, взятыми из спецификации шифра, представлены в первой колонке табл. 8. В остальных колонках этой таблицы представлены результаты зашифрования шифра Калина с комплектами (наборами) S-блоков, сгенерированных случайным образом. Хорошо видно, что свойства шифра для разных наборов S-блоков получились практически идентичными. Главное, что здесь есть свобода в выборе предпочтительного варианта. Заметим, что нулевое значение максимума дифференциальной таблицы связано с тем, что в экспериментах используется 16-битная разность на основе использования 4-го и 7-го байтов входа и выхода (байты нумеруются, начиная с нуля!). В результате после операции MixColumns первого цикла байты сдвигаются, и на местах 4-го и 7-го байтов выхода оказываются нулевые разности. Если бы формировались разности на основе, например, 0-го и 1-го байтов, то уже после первого цикла получили бы значение максимума дифференциальной таблицы равное 18 – 20 (сразу бы получили показатель максимума случайной подстановки).

В табл. 8 в первой колонке приведены поцикловые значения максимумов таблиц дифференциальных разностей для восьми стандартных подстановок шифра Калина. Во второй колонке приведены средние значения максимумов для 10 экспериментов с применением 80 случайных подстановок (по восемь на каждый эксперимент). В последней колонке представлены максимумы для первого набора из восьми случайных подстановок.

В табл. 9 по аналогии с шифром Rijndael представлены линейные показатели шифра Калина. И эти результаты свидетельствуют о том, что Калина со случайными S-блоками оказывается вполне конкурентоспособной шифру в авторской разработке.

S-блоки для шифра Мухомор

В шифре Мухомор S-блоки также приведены в спецификации без особых оговорок. В табл. 10 в первой колонке приведены результаты проверки дифференциальных показателей шифра Мухомор со стандартными S-блоками. Результаты тестирования шифров Мухомор со случайными S-блоками приведены в остальных колонках этой таблицы.

Таблица 8

Число циклов	Стандартные подстановки	Усредненное значение для 80 случайных подстановок (10 экспериментов)	Значение для восьми случайных подстановок (1 эксперимент)
1	20	0	0
2	18	25	28
3	18	19,6	20
4	20	18,6	18
5	20	18,8	18
6	20	19,4	20
7	18	18,6	20
8	18	19	18
9	18	19,2	18
10	20	19,6	18
11	20	19,2	20
12	20	19	18
13	18	19,4	20

Таблица 9

Число циклов	Стандартные подстановки	Усредненное значение для 80 случайных подстановок (10 экспериментов)	Значение для восьми случайных подстановок (1 эксперимент)
1	838	815,8	824
2	812	813	810
3	829	820,1	808
4	812	820,6	805
5	907	817,2	840
6	818	827,9	865
7	843	825	790
8	839	835,2	831
9	820	819,4	818
10	818	823	818
11	778	821	904
12	896	818	822
13	825	823,9	842

Таблица 10

Число циклов	Стандартные подстановки	Усредненное значение для 80 случайных подстановок (10 экспериментов)	Значение для четырех случайных подстановок (1 эксперимент)
1	18	19,8	20
2	18	19,4	20
3	18	19,2	20
4	20	18,6	18
5	20	19,8	18
6	18	19,6	20
7	18	18,8	20
8	18	19,6	20
9	20	19	20
10	20	18,8	18
11	20	19,8	22
12	20	19,2	22
13	18	18	20

В табл. 11 представлены линейные показатели этого шифра.

Таблица 11

Число циклов	Стандартные подстановки	Усредненное значение для 80 случайных подстановок (10 экспериментов)	Значение для четырех случайных подстановок (1 эксперимент)
1	833	819,3	813
2	843	830,3	861778
3	818	823,9	855800
4	817	815,8	831
5	835	821,3	804
6	895	810,3	819
7	833	826	835
8	799	812,4	834
9	837	823,8	795
10	817	830	796
11	847	825	802
12	830	839	813
13	833	803,5	816

Выводом из приведенных результатов является возможность использовать случайные S- блоки для реализации эффективного шифрующего преобразования.

S-блоки для шифра Лабиринт

В последней серии экспериментов представляем еще один из шифров прошедшего украинского конкурса – шифр Лабиринт.

И для него приводятся дифференциальные и линейные показатели при использовании в шифре родных и случайно взятых S-блоков.

Таблица 12

Число циклов	Стандартная подстановка	Усредненное значение для 10 случайных подстановок	Значение для одной случайно выбранной подстановки
1	20	101	104
2	20	19,6	18
3	18	19	18
4	20	18,6	20
5	18	19,6	18
6	20	19	18
7	20	19,4	20
8	20	18,6	20

В табл. 13 представлены линейные показатели шифра Лабиринт, которые свидетельствуют о том, что шифр Лабиринт со случайными S-блоками оказывается вполне конкурентоспособным шифру в авторской разработке.

Таблица 13

Число циклов	Стандартная подстановка	Усредненное значение для 10 случайных подстановок	Значение для одной случайно выбранной подстановки
1	839	832,8	826
2	835	831,4	840
3	826	828,3	881
4	832	817	808
5	880	815,8	851
6	829	774,2	798
7	827	830	851
8	786	828	843

Как и для предыдущих шифров, видим, что законы распределения обеих таблиц для случайно выбранных подстановок близки к аналогичным законам распределения для стандартных подстановок.

Выводы

Результатами работы подтверждено центральное положение новой методологии оценки стойкости блочных симметричных шифров к атакам линейного и дифференциального криптоанализа, состоящее в том, что показатели стойкости БСШ не зависят от используемых в шифрах S-блоков, а определяются показателями случайных подстановок соответствующей степени. Более того, показано, что в качестве S-блоков, позволяющих реализовать эффективное шифрующее преобразование, могут выступать S-блоки, выбранные случайным образом. Это значит, что задача поиска математических конструкций совершенных S-блоков, которой уделяется большое внимание в современной криптографической литературе, потеряла практический смысл.

Предложены конструкции случайных S-блоков для шифра Rijndael и шифров, представленных на украинский конкурс, которые по криптографическим показателям не уступают S-блокам, представленным в спецификациях шифров.

Получил дальнейшее развитие метод оценки эффективности оценки криптографических преобразований (итеративных шифров), основанный на оценке числа циклов, необходимых шифру для перехода к дифференциальным и линейным показателям, свойственным случайной подстановке.

Установлено, что по этому показателю криптографической пригодности шифры Калина, Мухомор и Лабиринт, представленные на украинский конкурс, являются более совершенными, чем шифр Rijndael. Они на 1-2 цикла имеют перед ним преимущество в динамике перехода к стационарным состояниям, свойственным случайным подстановкам соответствующей степени.

Список литературы: 1. *Lisitskaya, I.V.* Importance of S-Blocks in Modern Block Ciphers / I.V. Lisitskaya, A.A. Nastenko and K.E. Lysytskiy // Computer Network and Information Security. – 2012. – № 10, 1-12. 2. *Лисицкая, И.В.* Большие шифры – случайные подстановки / И.В. Лисицкая, А.А. Настенко // Радиотехника. – 2011. – Вып. 166. – С. 50-55. 3. *Лисицкая, И.В.* Дифференциальные свойства шифра FOX. / И.В. Лисицкая, Д. С. Кайдалов // Прикладная радиоэлектроника. – 2011. – Т. 10, № 2. – С. 122-126. 4. *Горбенко, І.Д.* Перспективний блоковий симетричний шифр "Калина" – основні положення та специфікації / І.Д. Горбенко, В.І. Долгов та ін. // Прикладна радіоелектроніка. – 2007. – Т.6. – № 2. – С. 195-208. 5. *Горбенко, І.Д.* Перспективний блоковий симетричний шифр «Мухомор» – основні положення та специфікація / І.Д. Горбенко, М.Ф. Бондаренко, В.І. Долгов та ін. // Прикладная радиоэлектроника. – 2007. – Т. 6, №2. – С. 147-157. 6. *Головашич, С.А.* Спецификация алгоритма блочного симметричного шифрования «Лабиринт» // Прикладная радиоэлектроника. – 2007. – Т. 6, №2. – С. 230-240. 7. *Олейников, Р.В.* Результаты анализа алгоритма шифрования ADE / Р.В. Олейников, В.И. Руженцев, М.С. Михайленко, А.Б. Небывайлов // Прикладная радиоэлектроника. – 2008. – Т. 7, № 3. – С. 210-214.

Харьковский национальный университет радиоэлектроники

Поступила в редколлегию 25.09.2012

УДК 621.3.06

А.В. КАЗИМИРОВ, Р.В. ОЛЕЙНИКОВ, канд. техн. наук

КРИПТОАНАЛИЗ ШИФРА MICKEY НА ОСНОВЕ АНАЛИЗА ВНУТРЕННИХ СОСТОЯНИЙ

Введение

Нелинейные регистры с обратной связью оказываются перспективными элементами для построения генераторов гаммы (поточных шифров). Такие криптопримитивы эффективны при аппаратной реализации и обеспечивают высокий уровень нелинейности, что имеет решающее значение для обеспечения криптографической стойкости. Тем не менее, работы [1, 2], посвященные анализу этих конструкций, основаны на ряде допущений и не имеют формальных оценок и доказательств. Ниже показано, что неудачный выбор параметров, которые, на первый взгляд, не влияют на стойкость, может кардинально снизить сложность криптоаналитической атаки.

Основными требованиями к гамме шифрующей являются большая длина периода, оптимальные корреляционные свойства и высокая нелинейность [3, 4]. Чтобы добиться этого, разработчики поточных шифров часто объединяют линейные и нелинейные регистры. При этом предполагается, что линейная часть гарантирует период гаммы шифрующей, а нелинейная обеспечивает стойкость к статистическим и алгебраическим атакам. Кроме того, функция обновления для обоих регистров может включать в себя состояние соседнего регистра и осуществлять так называемое взаимное управление. Современные генераторы гаммы, в дополнение к секретному ключу, часто используют открытый вектор инициализации (IV), что позволяет использовать один и тот же ключ в разных сеансах связи. Такой подход к построению был применен при разработке шифра Mickey [5, 6].

В предыдущей работе [7] были исследованы некоторые характеристики дерева обратных состояний алгоритма Mickey. В [8] представлена теоретическая оценка графа состояний. В представленной статье обобщаются предыдущие результаты и рассматриваются новые методы анализа шифра.

Краткое описание поточного шифра MICKEY

Существуют две версии шифра - Mickey-80 и Mickey-128 [5, 6]. Обе версии основаны на комбинации линейного (R) и нелинейного (S) регистров. Алгоритм имеет два входных параметра: вектор инициализации (IV) и сеансовый ключ (K). Общая структура алгоритма показана на рис. 1. Максимальная длина сообщения, которая может быть получена при использовании одной пары (K, IV), равна 2^{40}. Различия в параметрах между версиями представлены в табл. 1.

Ячейки регистров являются битовыми и обозначаются r_0, r_1, …, r_{RL-1} и s_0, s_1, …, s_{RL-1}. R и S тактируются в соответствии с управляющими битами $CB_R = CB_SL \oplus CB_RR$ для регистра R и $CB_S = CB_SR \oplus CB_RL$ для S – соответственно, где CB_XY означает определённый бит регистра X из табл. 1 (например, $CB_SL = s_{34}$ для Mickey-80). Функция обновления регистров имеет дополнительный параметр – входной бит (IB_S и IB_S).

Таблица 1

Различия между параметрами шифров Mickey-80 и Mickey-128

Версия	Длина регистра (RL)	Длина ключа	Длина холостого хода	*CB_SL*	*CB_SR*	*CB_SM*	*CB_RL*	*CB_RR*
80	100	80	100	34	67	50	33	67
128	160	128	160	54	106	80	53	106

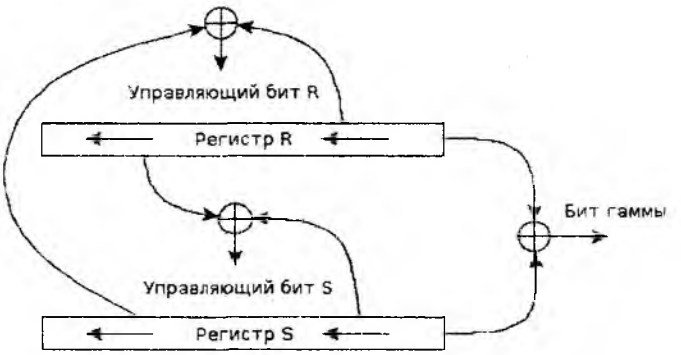


Рис. 1. Обобщённая структура шифра Mickey

Как видно из табл. 1, соотношение длины регистров ко всем остальным параметрам почти одинаково для обеих версий шифра. Несмотря на этот факт, дерево обратных состояний для Mickey-80 и Mickey-128 значительно отличается, что показано ниже.

Шифр имеет следующие режимы работы:

– инициализация регистров *R* и *S* нулями;
– загрузка IV (*CLOCK_K_IV*);
– загрузка ключа (*CLOCK_K_IV*);
– холостой ход (*CLOCK_PRECLOCK*);
– генерация гаммы (*CLOCK_KG*).

Все режимы, кроме генерации гаммы, также работают в так называемом режиме смешивания. Это означает, что входной бит регистра *R* (*IB_R*) зависит от бита *CB_SM* регистра *S*. Более подробное описание приведено в работе [7].

Модель анализа нелинейного генератора гаммы шифрующей

В [9] уже была описана модель атаки, основанная на знании внутреннего состояния регистров. Далее приводится её краткое описание и применение к шифру Mickey.

Предполагается, что злоумышленник знает внутреннее состояние генератора гаммы (состояние регистров *R* и *S*) и сколько тактов было выполнено, чтобы попасть в это состояние из начального. Оценка стойкости шифра Mickey основана на построении дерева обратных состояний, которое рассчитывается с использованием обратных функций тактирования шифра $CLOCK_PRECLOCK^{-1}$, $CLOCK_K_IV^{-1}$ и $CLOCK_KG^{-1}$ (примеры алгоритмов приведены в [7]). Получение обратных состояний сводится к полному перебору всех возможных значений входных параметров (бита обратной связи, управляющего и входного битов) и исключение невозможных состояний.

Стоит отметить, что предыдущее состояние не всегда определяется однозначно. Количество ветвлений может варьироваться в зависимости от известного криптоаналитику состояния и режима шифрования. В зависимости от режима, можно получить три различных дерева состояний. Любое из этих деревьев может рассматриваться как граф переходов конечного автомата. В общем случае дерево показано на рис. 2. В дальнейшем будут использоваться понятия, связанные с деревом, такие как уровень - множество обратных состояний, определяющие исходное состояние после определенного количества тактов и степень ветвления (СВ) - число возможных предыдущих состояний, которые дают состояние *R* и *S* за один такт вперед. Точки ветвления в дереве обратных состояний хранят значение степени ветвления. В режиме загрузки ключа/IV рёбра помечаются соответствующими искомыми битами ключа. Используя приведенную структуру дерева, при достижении состояния со СВ = 0 можно отсечь часть дерева, что приводит к уменьшениюпространственной сложности атаки.

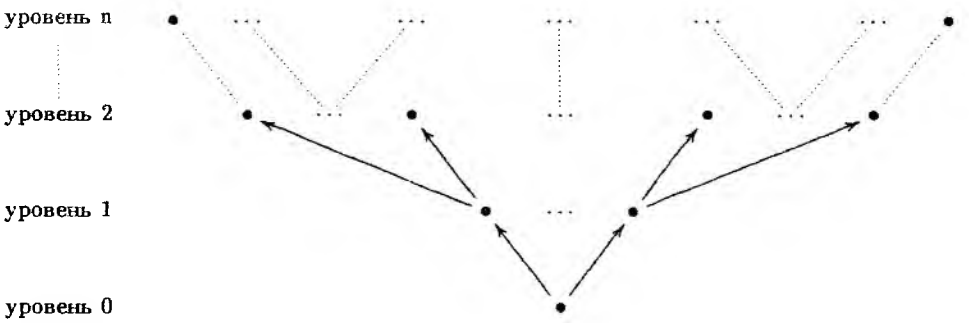


Рис. 2. Дерево обратных состояний

Предыдущая модель может быть сведена к более частному случаю в предположении, что злоумышленник знает внутреннее состояние генератора гаммы после определенного числа тактов в режиме загрузки ключа. Это количество тактов должно быть небольшим для отката состояния генератора до начала загрузки ключа (конец загрузки IV). Знание IV злоумышленником приводит к тому, что большая часть состояния дерева может быть удалена. На основании этого можно практически однозначно определить ключ, так как начальное состояние генератора фиксировано, и только один путь в дереве будет соответствовать реальному обновлению состояний регистра.

Другой способ нахождения ключа основывается на атаке типа "встреча посередине". Пусть известно состояние регистра после k-тактов от начала режима загрузки ключа, что так же соответствует k битам ключа. Схема атаки состоит из трёх этапов и представлена на рис. 3.

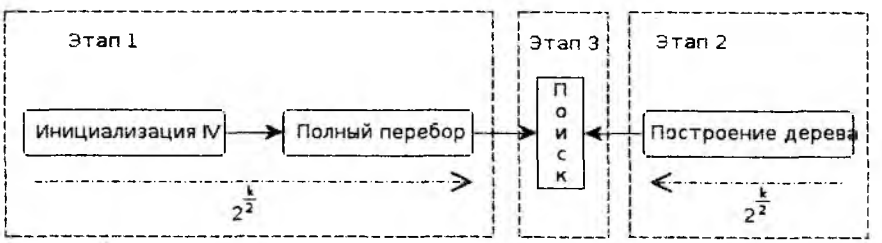


Рис. 3. Атака типа "встреча посередине" на шифр Mickey

На первом этапе для известного IV вычисляется состояние, которое эквивалентно входному значению в функцию инициализации ключа. Для полученного значения находятся все возможные состояния для $\frac{k}{2}$ битов ключа с применением функции *CLOCK_K_IV*. На втором этапе, на основе внутреннего состояния регистров, полученных после k тактов генератора, с использованием функции $CLOCK_K_IV^{-1}$ строится дерево обратных состояний вплоть до $\frac{k}{2}$ уровня. После этого, на последнем этапе находятся одинаковые состояния во множествах, сформированных на предыдущих этапах.

Сложность атаки полного перебора или построения дерева обратных состояний приблизительно равна 2^k. Однако комбинация этих двух методов в атаке типа "встреча посередине" является более практичной и даёт сложность

$$O_d(k) + O_i(k) + O_f(k) = O_d(2^{\frac{k}{2}}) + O_i(2^{\frac{k}{2}}) + O_f(2^{\frac{k}{2}}) \approx 2^{\frac{k}{2}+2},$$

где $O_d(k)$ – сложность атаки полного перебора для $\frac{k}{2}$ битов ключа, $O_i(k)$ – сложность построения дерева обратных состояний и $O_f(k)$ – сложность нахождения одинаковых состояний; используется хэш-таблица [10].

Результаты анализа шифра MICKEY

Количество допустимых предшествующих состояний

Как отмечено выше, предполагается, что криптоаналитик знает состояние регистров после определённого количества тактов работы шифра.

Вероятность получения степени ветвлений предшествующих состояний, определенная на основе анализа дерева обратных состояний на глубину до 18 уровней (обратных тактов) с помощью функции $CLOCK_X^{-1}$, приведена в табл. 2 (случайное заполнение внутреннего состояния регистров) и табл. 3 (верное определённое состояние).

Таблица 2

Вероятность степени ветвления для случайного начального состояния

СВ	Загрузка ключа/IV		Холостой ход		Генерация гаммы	
	80	128	80	128	80	128
0	0.2982	0.198	0.2802	0.2825	0.3014	0.2718
1	0.00009	0.1031	0.4377	0.4590	0.4052	0.4281
2	0.4229	0.4022	0.2735	0.2294	0.2844	0.29
3	0.0001	0.1087	-	0.0256	-	-
4	0.2698	0.1703	0.0085	0.0035	0.0090	0.0101
6	0.00001	0.0177	-	-	-	-
8	0.0089	-	-	-	-	-

Таблица 3

Вероятность степени ветвления для верного определённого начального состояния

СВ	Загрузка ключа/IV		Холостой ход		Генерация гаммы	
	80	128	80	128	80	128
0	0.2773	0.2186	0.3052	0.29	0.3041	0.3038
1	0.00001	0.1047	0.4345	0.4534	0.4323	0.4154
2	0.4331	0.3753	0.2523	0.2256	0.2558	0.2698
3	0.00002	0.1029	-	0.0289	-	-
4	0.28	0.1783	0.008	0.0021	0.0079	0.0111
6	0.00007	0.0203	-	-	-	-
8	0.0095	-	-	-	-	-

Как видно из табл. 2 и 3, вероятности для случайного и верного определённого заполнения регистров примерно одинаковы и соответствуют теоретическим (вычисленным по методике, приведенной в [8]).

Математическое ожидание числа степени ветвлений для режима загрузки ключа/IV приблизительно равно 2, в то время как для режима холостого хода и генерации гаммы - 1. Таким образом, получив значение регистров на любом из этапов, криптоаналитик может выполнить обратное тактирование и получить состояние регистров после инициализации ключа. Сложность этапа приблизительно равна $O(k)$, где k – количество тактов, сделанных регистром после завершения этапа инициализации ключа.

В табл. 4 - 6 показано среднее число допустимых состояний на каждом уровне для режимов загрузки ключа/IV, холостого хода и генерации гаммы соответственно. Усреднение проводилось на 1000 случайных состояний, чтобы исключить зависимость от начального состояния.

Таблица 4

Количество допустимых предшествующих состояний
в зависимости от уровня дерева обратных состояний в режиме загрузки ключа/IV

Уровень	Количество состояний	
	80	128
0	1	1
1	3	3
2	9	7
3	25	18
4	45	39
5	143	82
6	247	171
7	523	347
8	1183	703
9	2221	1435
10	5075	2904
11	9453	5849
12	18694	11834
13	37702	23801

Таблица 5

Режим холостого хода

Уровень	Количество состояний	
	80	128
0	1	1
1	1	1
2	1	2
3	2	3
4	3	4
5	4	6
6	5	8
7	7	8
8	5	11
9	5	14
10	10	15
11	12	15
12	13	15
...	...	...
99	25	30
100	32	27
...	...	...
159	-	59

Таблица 6

Режим генерации гаммы

Уровень	Количество состояний	
	80	128
0	1	1
1	1	2
2	1	3
3	2	4
4	3	6
5	3	6
6	4	4
7	5	3
8	3	4
...	...	...
125	17	64
126	17	81
127	16	91
128	20	97

Как видно из табл. 4 - 7, количество допустимых предшествующих состояний увеличивается в соответствии с математическим ожиданием индекса степени ветвлений для соответствующего режима.

Определение битов ключа на основе дерева обратных состояний

Дерево обратных состояний, описанное выше, позволяет криптоаналитику вычислять значение некоторых битов ключа. Их количество и расположение определяется структурой дерева и зависит от конкретного значения вектора инициализации и ключа шифрования (табл. 7).

С увеличением уровня увеличивается и вероятность появления тупикового состояния (для которого отсутствуют допустимые предшествующие). Каждый обратный шаг увеличивает вероятность отсечения поддерева с множеством неверных значений битов ключа.

Как видно из колонки для Mickey-128, биты ключа могут быть однозначно определены на уровнях 1 и 3 ("1" и "0" соответственно). Однако, зная эти данные, невозможно найти однозначно обратное состояние на этих уровнях, следовательно, они не уменьшают дерево и сложность определения всех битов ключа. Более того, возможность нахождения бита ключа напрямую зависит от начального состояния. Например, для Mickey-80 невозможно определить биты ключа с вероятностью, равной 1.

Таблица 7

Распределение вероятностей значений битов ключа на пяти уровнях дерева

Уровень	Вероятность			
	80		128	
	1	0	1	0
1	0.5	0.5	1	0
2	0.5	0.5	0.5	0.5
3	0.5	0.5	0	1
4	0.5	0.5	0.5	0.5
5	0.4857	0.5143	0.5	0.5

Нахождение одинаковой гаммы для различных пар ключа и IV

Функции, используемые в различных режимах шифра Mickey, допускают генерацию гаммы шифрующей, сдвинутой на несколько бит для различных пар ключа и IV (перекрытие гаммы).

Пусть z_i^h – i-й бит гаммы для h-го набора (K_h, IV_h), где ключ имеет вид $K_1 = \{k_0, k_1, \ldots, k_{n-1}\}$, а n – длина ключа в битах (табл. 1). Тогда можно найти такие K_2, IV_1 (значение первого вектора инициализации) и IV_2, для которых состояние регистров будут отличаться на один шаг, а выходная гамма иметь свойство $z^2_i = z^1_{i+1}$.

Допустим, что $IV_1 = \{iv_0, iv_1, \ldots, iv_j\}$, $IV_2 = \{iv_0, iv_1, \ldots, iv_j, k_0\}$ и $K_2 = K_1 << 1 = \{k_1, k_2, \ldots, k_{n-1}, 0\}$. Относительное размещение битов для различных наборов ключа/IV приведены в табл. 8.

Таблица 8

Отличия параметров для различных наборов ключа и IV

IV_1				K_1						Холостой ход				Генерация гаммы					
iv_0	iv_1	…	iv_j	k_0	k_1	k_2	…	k_{n-2}	k_{n-1}	0	0	0	…	0	0	0	0	0	…
iv_0	iv_1	…	iv_j	k_0	k_1	k_2	…	k_{n-2}	k_{n-1}	0	0	0	…	0	0	0	0	0	…
IV_2				K_2						Холостой ход				Генерация гаммы					

Очевидно, что состояния регистров отличаются на один шаг, что, в свою очередь, приводит к формированию одинаковой гаммы шифрования, сдвинутой на один бит. Это обусловлено тем, что режим холостого хода эквивалентен режиму инициализации ключа/IV, когда "входной бит" равен 0.

Отличия состояния регистров на один шаг означает, что гамма шифрования будет иметь те же свойства, то есть $z^2_i = z^1_{i+1}$. Перекрытие в шифре Mickey присутствует из-за того, что режим холостого хода эквивалентен режиму инициализации ключа/IV, когда "входной бит" равен 0. Должно выполняться дополнительное условие: $s^1_{50} = 0$ в момент перехода между режимами холостого хода и генерации гаммы. Более того, длина IV_2 может быть увеличена на определенное количество бит, и в результате гамма шифрования будет сдвинута на такое же значение (при условии соответствия ключа K_2 вектору инициализации, см.табл. 8).

Более того, длина IV_2 может быть изменена на другое количество бит, в результате K_2 и гамма шифрования будут сдвинуты на такое же значение.

Аналогичные рассуждения могут быть применены к 128-разрядной версии Mickey.

Ниже представлены примеры одинаковых гамм для различной длины IV. Все значения имеют шестнадцатеричное представление, кроме первого значения Z_1 и последнего Z_2 – они являются битами.

Значения для Mickey-80:

K_1 = { d3, ec, f0, 84, 8a, 1d, b1, b7, 4a, dd }
IV_1 = { 58, e5, 77, 0a, 9c, a2, 34, c7, cd, 5e } (79 bits)
K_2 = { a7, d9, e1, 09, 14, 3b, 63, 6e, 95, ba }
IV_2 = { 58, e5, 77, 0a, 9c, a2, 34, c7, cd, 5f } (80 bits)
Z_1 = { 0, B7, 61, 27, 92, C5, 85, 91, 51, 18, 2A, D6, 7C, 8C, C8, C7, 04 }
Z_2 = { B7, 61, 27, 92, C5, 85, 91, 51, 18, 2A, D6, 7C, 8C, C8, C7, 04, 1 }

Значения для Mickey-128:

K_1 = { c9, 55, e7, 7a, 80, 13, 1a, ad, 40, 45, d9, 6c, 71, 04, 97, 9c }
IV_1 = { 4e, db, 6e, 01, 05, 98, 2b, 30, c3, 56, 5a, ed, 80, 85, 18, aa } (127 bits)
K_2 = { 92, ab, ce, f5, 00, 26, 35, 5a, 80, 8b, b2, d8, e2, 09, 2f, 38 }
IV_2 = { 4e, db, 6e, 01, 05, 98, 2b, 30, c3, 56, 5a, ed, 80, 85, 18, ab } (128 bits)
Z_1 = { 0, 79, 23, 91, 05, E1, DD, 2D, 9D, 83, 3E, B4, 78, 52, E5, A6, 66 }
Z_2 = { 79, 23, 91, 05, E1, DD, 2D, 9D, 83, 3E, B4, 78, 52, E5, A6, 66, 1 }

Выводы

Шифр Mickey является генератором гаммы, основанном на одновременном использовании линейного и нелинейного регистров. Проведенный анализ алгоритма показал, что криптоанализ на основе обратного тактирования возможен во всех режимах, включая режим загрузки ключа/IV. Несмотря на то, что переходы между внутренними состояниями осуществляются псевдослучайным образом, в некоторых случаях, возможно восстановление битов ключа шифрования.

Особенности функций тактирования дают возможность находить пары ключа и IV, использование которых приводит к сдвигу гамма шифрования на определенное количество позиций. Однако ограничение, накладываемое в спецификации шифра на длину вектора инициализации при использовании одного ключа, не позволяет реализовать практическую атаку.

Корректность известной теоретической методики [8] для оценки вероятности степени ветвлений шифра Mickey дополнительно подтверждается нашими практическими результатами, что позволяет обосновать выбор параметров Mickey-подобных алгоритмов шифрования на стадии проектирования.

Список литературы: 1. *Dubrova, E., Maxim,T., Hannu, T.* On Analysis and Synthesis of (n,k)-Non-Linear Feedback Shift Registers. DATE 2008. – pp. 1286-1291. 2. *Dubrova, E.* A List of Maximum Period NLFSRs. Electronic resource. Available at http://eprint.iacr.org/2012/166.pdf. 3. *Rueppel, R.* Analysis and Design of Stream Ciphers. - Springer-Verlag, 1986. - 244 p. 4. *Грушо, А. А.* Теоретические основы компьютерной безопасности : учеб. пособие для студентов высш. учеб. заведений / А. А. Грушо, Э. А. Применко, Е. Е. Тимонина. – М. : Изд. центр «Академия», 2009. – 272 с. 5. *Babbage, S., Dodd, M.* The Stream Cipher Mickey 2.0. eSTREAM, ECRYPT Stream Cipher Project, End of 3rd Phase, September 2009. http://www.ecrypt.eu.org/stream/ p3ciphers/mickey/mickey_p3.pdf. 6. *Babbage, S., Dod, M.* The stream cipher MICKEY-128 2.0. eSTREAM, ECRYPT Stream Cipher Project, End of 3rd Phase, September 2009. http://www.ecrypt.eu.org/ stream/p3ciphers/mickey/mickey128_p3.pdf. 7. *Казимиров, А.В., Олейников, Р.В.* Оценка количества допустимых внутренних состояний в поточном алгоритме Mcikey // Прикладная радиоэлектроника. – 2011. – Т.10, №2. - С. 112-115. 8. *Jansen, C.J.A.* The State Space Structure of the Mickey Stream Cipher. In Horlin, F., ed. // 32nd Symposium on Information Theory in the Benelux, Brussels, Werkgemeenschap voor Informatien Communicatietheorie (2011). 9. *Golic, J. D.* Cryptanalysis of alleged A5 stream cipher, in Advances in Cryptology – EUROCRYPT'97 (W. Fumy, ed.), vol. 1233 of Lecture Notes in Computer Science, Springer-Verlag, 1997. – P. 239-255. 10 *Кормен, Т., Лейзерсон, Ч., Ривест, Р., Штайн, К.* Хеш-таблицы. Гл. 11. // Алгоритмы: построение и анализ ; под ред. И. В. Красикова. – 2-е изд. – М. : Вильямс, 2005. – 1296 с.

Харьковский национальный университет радиоэлектроники

Поступила в редколлегию 21.09.2012

УДК 621. 3.06

И.В. ЛИСИЦКАЯ, д-р техн. наук

ВЫРОЖДЕННЫЕ ПОДСТАНОВКИ

Введение

Одним из основных положений, развиваемых в новой методологии оценки показателей стойкости блочных симметричных шифров (БСШ) к атакам дифференциального и линейного криптоанализа, является положение, состоящее в том, что все БСШ после определенного индивидуального для каждого БСШ числа циклов приходят по дифференциальным и линейным показателям к свойствам случайных подстановок соответствующей степени. В концентрированном виде эти идеи выражены в работе [1] следующим образом.

Все современные блочные симметричные шифры через определенное число циклов независимо от используемых в шифрах S блоков (здесь идет речь не о вырожденных их конструкциях) приходят к свойствам случайных подстановок, то есть по комбинаторным показателям (числу инверсий, циклов и возрастаний), а также по законам распределения переходов XOR таблиц (таблиц полных дифференциалов) и законам распределения смещений таблиц линейных аппроксимаций (линейных корпусов) повторяют соответствующие показатели случайных подстановок. Вследствие этого значения максимумов полных дифференциалов и линейных корпусов могут быть получены расчетным путем из формул для законов распределения вероятностей переходов XOR таблиц и смещений таблиц линейных аппроксимаций случайных подстановок.

При этом проверка показателей случайности больших шифров может быть выполнена на основе разработки и последующего анализа показателей случайности уменьшенных моделей, которые допускают проведение вычислительных экспериментов в реальные временные сроки.

В качестве возражений выдвинутому положению существует мнение о том, что это всё частные результаты, и можно без труда найти подстановки, которые не укладываются в его рамки.

Проведенные многочисленные эксперименты свидетельствуют, однако, о том, что сформулированное положение выполняется практически для всех известных шифров, а подстановки, его не подтверждающие, – это ограниченное множество по сравнению с общим числом возможных подстановок, существенно не влияющее на достоверность сформулированного утверждения.

Из сказанного следует, что предлагаемая методология работает для произвольных S-блоковых конструкций, исключая вырожденные подстановки. Возникает вопрос, а какие же S-блоки считать вырожденными ?

Одним из простых ответов на этот вопрос может стать следующее: вырожденными S-блоками следует считать те, которые либо не позволяют шифру в пределах ограниченного числа циклов, однозначно определенного для каждого шифра, достичь показателей случайной подстановки, либо стационарное значение, к которому приходит шифр, не соответствует ожидаемому, свойственному случайной подстановке.

Эта работа посвящена изучению представителей множества вырожденных подстановок, а также оценке мощности этого множества, что позволит удостовериться в выполнении сформулированного положения с весьма высоким уровнем доверия.

Примеры вырожденных подстановок

Здесь представляются результаты экспериментов с малыми (16-битными) моделями шифров. Подтверждение адекватности малых моделей своим большим прототипам можно найти в [2, 3 и др.]. Соответственно разговор будет идти о полубайтовых S-блоках, изученных наиболее всесторонне [4, 5 и др.].

Эксперименты показывают, что к вырожденным подстановкам следует отнести, прежде всего, подстановки с показателями нелинейности, равными или близкими к нулю (максимальное значение смещения таблиц ЛАТ равно 2^{n-1}). В табл. 1 и 2 представлено поцикловое поведение полных дифференциалов 16-битного шифра Хеуса из работы [6] (шифра со слабым линейным преобразованием).

Таблица 1

Подстановка		Значения максимумов ТР в зависимости от числа циклов					
Тождественная подстановка		Число циклов					
1	1,2,3,4,5,6,7,8,9,1A,B,C,D,E,F ЛАТ – 8, ДТ – 16 (15)	1	2	3	4	5	6
		57617,07	50364,40	45675,60	40971,40	37338,80	39267,00
		Число циклов					
		7	8	9	10	11	12
		41487,80	43386,53	44803,13	46411,40	47075,40	47872,93
Подстановка 14 из работы []		Число циклов					
2	5,A,C,6,0,F,3,9,8,D,B,1,7,2,E,4 ЛАТ – 8, ДТ – 8 (12)	1	2	3	4	5	6
		32768	16384	5043,20	1327,87	369,60	151,07
		Число циклов					
		7	8	9	10	11	12
		61,53	32,60	24,20	23,87	23,93	24,13
Подстановка 1 из работы [22]		Число циклов					
3	C,D,5,1,A,B,6,2,E,3,7,F,4,0,8,9 ЛАТ – 8, ДТ – 12 (3)	1	2	3	4	5	6
		49152,00	27648,00	15552,00	3616,00	1016,00	451,27
		Число циклов					
		7	8	9	10	11	12
		209,27	106,60	53,07	27,53	20,07	19,07

В качестве первого примера взята тождественная подстановка (единичная подстановка симметрической группы). Результаты её применения для построения процедуры зашифрования (шифра) приведены в верхней части табл. 1 (первый пример). Этот пример свидетельствует, что без применения подстановочной нелинейной операции шифр (любой) просто разваливается (он далёк от случайной подстановки).

Второй и третий примеры подстановок взяты из работ [5, 7]. Они свидетельствуют, что действительно существуют и нетождественные подстановки со значением показателя максимума смещения линейной аппроксимационной таблицы (ЛАТ), равным 8 (максимально возможным значением для полубайтовой подстановки), которые также не позволяют реализовать эффективную процедуру зашифрования. Заметим, что вторая подстановка приходит к асимптотическому значению максимума таблицы дифференциалов (ТД) равному 24, отличающемуся от теоретического значения максимума дифференциала для случайной подстановки (18 – 20).

В табл. 2 представлены результаты поцикловой оценки максимумов смещений таблиц линейных аппроксимаций (линейных оболочек) шифра Хеуса с этими же подстановками. И в этом случае результаты свидетельствуют о практической непригодности рассмотренных первых двух подстановок для построения шифрующих преобразований. Последняя подстановка приходит после семи циклов к показателям случайной подстановки, однако для достижения необходимых дифференциальных свойств (см. табл. 1) ей требуется более 11 циклов зашифрования. В то же время здесь можно сослаться на результаты работ [7, 8 и др.], из которых следует, что случайно взятые подстановки с большой вероятностью приводят к эффективному шифрующему преобразованию (число циклов, необходимое для перехода к случайной подстановке для шифра Хеуса, не превышает шести).

Таблица 2

Подстановка		Значения максимумов ЛАТ в зависимости от числа циклов					
Тождественная подстановка		Число циклов					
1	0,1,2,3,4,5,6,7,8,9,1A,B,C,D,E,F ЛАТ – 8, ДТ – 16 (15)	1	2	3	4	5	6
		32768	32768	32768	32768	32768	32768
		Число циклов					
		7	8	9	10	11	12
		32768	32768	32768	32768	32768	32768
Подстановка 14 из работы [5]		Число циклов					
2	5,A,C,6,0,F,3,9,8,D,B,1,7,2,E,4	1	2	3	4	5	6
		32768	32768	32768	32768	32768	32768
		Число циклов					
		7	8	9	10	11	12
		32768	32768	32768	32768	32768	32768
Подстановка 1 из работы [8]		Число циклов					
3	C,D,5,1,A,B,6,2,E,3,7,F,4,0,8,9	1	2	3	4	5	6
		32768	24576	12288	5244	2044	1080
		Число циклов					
		7	8	9	10	11	12
		792	872	826	816	842	816

Таблица 3

Подстановка		Значения максимумов ТР в зависимости от числа циклов					
Тождественная подстановка		Число циклов					
1	1,2,3,4,5,6,7,8,9,1A,B,C,D,E,F ЛАТ – 8, ДТ – 16 (15)	1	2	3	4	5	6
		57617,07	50364,40	45675,60	40971,40	37338,80	39267,00
		Число циклов					
		7	8	9	10	11	12
		41487,80	43386,53	44803,13	46411,40	47075,40	47872,93
Подстановка 14 из табл. 2		Число циклов					
2	5,A,C,6,0,F,3,9,8,D,B,1,7,2,E,4 ЛАТ – 8, ДТ – 8 (12)	1	2	3	4	5	6
		32768	16384	5043,20	1327,87	369,60	151,07
		Число циклов					
		7	8	9	10	11	12
		61,53	32,60	24,20	23,87	23,93	24,13
Подстановка 1 из работы [22]		Число циклов					
3	C,D,5,1,A,B,6,2,E,3,7,F,4,0,8,9 ЛАТ – 8, ДТ – 12 (3)	1	2	3	4	5	6
		49152,00	27648,00	15552,00	3616,00	1016,00	451,27
		Число циклов					
		7	8	9	10	11	12
		209,27	106,60	53,07	27,53	20,07	19,07

В табл. 3 приведены поцикловые распределения максимумов числа переходов XOR таблиц для шифров с сильным линейным преобразованием. Они реализуются с помощью уменьшенной (16-битной) конструкции шифра baby-Rijndael, в котором используются те же S-блоки, что и в предыдущих экспериментах. Первая вырожденная подстановка (тождественная) конечно и в этом случае приводит к развалу процедуры шифрования. В то же время другие две подстановки (по крайней мере, последняя на 11-ти циклах выходит к свойствам случайной подстановки). В то же время по линейным свойствам (см. табл. 4) вторая подстановка повторяет показатели, продемонстрированные шифром Хеуса (см. табл. 3). Она явно не подходит для построения шифра.

Далее было изучено влияние на качество шифрующего преобразования и дифференциальных свойств S-блоков. В табл. 5 приведены примеры подстановок с предельными (по максимуму) дифференциальными показателями. По результатам табл. 5 можно сделать

вывод, что плохими могут быть подстановки и со значением максимумов дифференциальных таблиц превыщающих 8-10 (они, как правило, свойственны подстановкам с максимальными значениями смещения ЛАТ).

Вырожденными могут быть подстановки и с немаксимальными значениями дифференциальных и линейных показателей. Пример такой подстановки представлен в табл. 6.

Таблица 4

Подстановки		Значения максимумов ЛАТ в зависимости от числа циклов					
Тождественная подстановка		Число циклов					
1	1,2,3,4,5,6,7,8,9,1A,B,C,D,E,F	1	2	3	4	5	6
		32768	32768	32768	32768	32768	32768
		Число циклов					
		7	8	9	10	11	12
		32768	32768	32768	32768	32768	32768
Подстановка 14 из табл. 2		Число циклов					
2	5,A,C,6,0,F,3,9,8,D,B,1,7,2,E,4	1	2	3	4	5	6
		32768	32768	32768	32768	32768	32768
		Число циклов					
		7	8	9	10	11	12
		32768	32768	32768	32768	32768	32768
Подстановка 1 из работы [22]		Число циклов					
3	C,D,5,1,A,B,6,2,E,3,7,F,4,0,8,9 ЛАТ – 8, ДТ – 12 (3)	1	2	3	4	5	6
		32768	24576	12288	5244	2044	1080
		Число циклов					
		7	8	9	10	11	12
		792	872	826	816	842	816

Таблица 5

Подстановка с тремя переходами с вероятностью 1		Число циклов					
4	F,D,9,B,1,3,2,0,4,6,7,5,A,8,C,E ЛАТ – 8, ДТ – 16 (3)	1	2	3	4	5	6
		58112,00	16827,00	8192,00	8192,27	8192,47	8193,93
		Число циклов					
		7	8	9	10	11	12
		8195,87	8195,47	8197,47	8198,20	8198,47	8200,87
Подстановка с одним переходом с вероятностью 1		Число циклов					
5	6,4,1,3,0,2,7,5,E,C,D,F,8,A,9,B ЛАТ – 8, ДТ – 16 (1)	1	2	3	4	5	6
		57617,07	10290,13	8192,00	8193,67	8192,47	8196,93
		Число циклов					
		7	8	9	10	11	12
		8198,67	8199,27	8200,27	8202,93	8204,13	8207,27

Таблица 6

Подстановка		Число циклов					
6	C,9,4,6,8,E,D,5,3,F,B,0,A,2,1,7 ЛАТ – 6, ДТ – 8 (5)	1	2	3	4	5	6
		32768,00	1536,00	139,07	23,53	24,20	23,73
		Число циклов					
		7	8	9	10	11	12
		23,80	24,07	23,80	23,93	24,00	23,80

В этом случае шифр пришёл к другому стационарному значению равному 24. Это второй пример подстановки с таким свойством (см. табл. 1). Заметим, что при другой конструкции линейного преобразования (линейным преобразованием MixColumn и ShiftRows GF(2^8)) шифр приходит к асимптотическому значению, соответствующему случайной подстановке. Нам пока не удалось объяснить этот эффект. Табл.7 демонстрирует линейные показатели шифра baby-Rijndael с этим же S-блоком (S-блоками).

Таблица 7

Подстановка		Значения максимумов ЛАТ в зависимости от числа циклов					
		Число циклов					
6	C,9,4,6,8,E,D,5,3,F,B,0,A,2,1,7 ЛАТ – 6, ДТ – 8 (5)	1	2	3	4	5	6
		32768,00	32768,00	32768,00	32768,00	32768,00	32768,00
		Число циклов					
		7	8	9	10	11	12
		32768,00	32768,00	32768,00	32768,00	32768,00	32768,00

Результат говорит о том, что этот S-блок нельзя применять для построения шифра.

Оценка доли вырожденных подстановок среди подстановок симметрической группы

Здесь покажем, что вероятность попасть на вырожденную подстановку при их случайном формировании весьма мала. Начнём с того, что вероятность получения подстановки с показателем нелинейности равным нулю, можно определить из формулы для соответствующего закона распределения вероятностей из работы [9]. В соответствии с этой работой для случайной подстановки степени 2^n справедлива теорема:

Теорема 1. *Пусть* $\lambda^*(\alpha,\beta)$ *будет случайным значением смещения линейной аппроксимационной таблицы* $LAT_\pi^*(\alpha,\beta)$ *для пары её входов* α *и* β*, когда подстановка* π *выбрана равновероятно из множества* 2^n *и* α,β *не нулевые. Тогда смещения* $\lambda^*(\alpha,\beta)$ *принимают только четные значения и для* $|l| \le 2^{n-2}$

$$\Pr\left(\lambda^*(\alpha,\beta)=|2l|\right)=\frac{\left(2^{n-1}!\right)^2}{2^n!}\cdot\begin{pmatrix}2^{n-1}\\ 2^{n-2}-|l|\end{pmatrix}^2. \qquad (1)$$

При $l = 2^{n-2}$ (показатель нелинейности 0) из этого соотношения получаем

$$\Pr\left(\lambda^*(\alpha,\beta)=2\cdot 2^{n-2}\right)=\frac{\left(2^{n-1}!\right)^2}{2^n!}\cdot\begin{pmatrix}2^{n-1}\\ 2^{n-2}-2^{n-2}\end{pmatrix}^2=$$

$$=\frac{\left(2^{n-1}!\right)^2}{2^n!}=\begin{pmatrix}2^n\\ 2^{n-1}\end{pmatrix}^{-1}\approx\left(2^{2^n}\frac{2^{-(n-2)/2}}{\sqrt{2\pi}}\right)^{-1}.$$

Здесь использована аппроксимация биномиального коэффициента из работы [10].

Из полученного результата следует, что для вероятности получить путём случайного выбора полубайтовую подстановку ($n = 4$) со значением максимума смещения таблицы линейных аппроксимаций равным $k = 2^3 = 8$ имеем

$$\Pr\left(\lambda(\alpha,\beta)=2\cdot 2^3\right)=7{,}65\cdot 10^{-5},$$

т.е. как показывают расчёты, эта вероятность меньше одной десятитысячной.

Эти расчёты сделаны для вероятности соответствующего заполнения одной ячейки таблицы, а таких ячеек в таблице (исключая нулевые маски по входу и выходу) $(2^n-1)^2$. Поэтому для вероятности получить полубайтовую подстановку со значением смещения $k = 2^3 = 8$ приходим к результату

$$7{,}65\cdot 10^{-5}\cdot(2^4-1)^2 = 0{,}0172.$$

Для генерации байтовой подстановки с таким же (нулевым) показателем нелинейности соответственно приходим к вероятности

$$\Pr\left(\lambda(\alpha,\beta) = 2\cdot 2^7\right) = 1{,}73\cdot 10^{-76}.$$

С учётом числа ячеек таблицы

$$1{,}73\cdot 10^{-76}\cdot(2^8-1)^2 = 1{,}13\cdot 10^{-71},$$

т.е. получить путём отбора такую подстановку практически невозможно.

Аналогично можно выполнить оценку вероятности порождения случайной подстановки с теоретически максимально возможным значением перехода XOR таблицы подстановки. Для этого воспользуемся теоремой из работы [11]. Она определяет вероятность события, заключающегося в том, что значение дифференциальной таблицы случайно взятой подстановки π порядка 2^n для перехода входной разности ΔX в соответствующую выходную разность ΔY будет равно $2k$. Эта вероятность в работе обозначена $\Pr(\Lambda_\pi(\Delta X, \Lambda Y) = 2k)$. Сама теорема звучит так:

Теорема 2. *Для любых ненулевых фиксированных* $\Delta X, \Delta Y \in Z_2^n$ *в предположении, что подстановка* π *выбрана равновероятно из множества* S_2^n *и* $0 \le k \le 2^{n-1}$

$$\Pr(\Lambda_\pi(\Delta X, \Lambda Y) = 2k) = \binom{2^{n-1}}{k}^2 \cdot \frac{k!2^k\cdot\Phi(2^{n-1}-k)}{2^n!}, \qquad (2)$$

где функция $\Phi(d)$ *определяется выражением*

$$\Phi(d) = \sum_{i=0}^{d}(-1)^i\cdot\binom{d}{i}^2\cdot 2^i\cdot i!(2d-2i)!$$

Здесь для $k = 2^{n-1}$ получим расчётное соотношение

$$\Pr(\Lambda_\pi(\Delta X, \Lambda Y) = 2\cdot 2^{n-1}) =$$

$$= \binom{2^{n-1}}{2^{n-1}}^2 \frac{2^{n-1}!2^{2^{n-1}}\cdot\Phi(2^{n-1}-2^{n-1})}{2^n!}.$$

Поскольку, как отмечено в [10], $\Phi(d)\approx(2d)!e^{-\frac{1}{2}}$, то $\Phi(2^{n-1}-2^{n-1}) = \Phi(0)\approx e^{-\frac{1}{2}}$, тогда:

$$\Pr\left(\Lambda(\Delta X, \Delta Y) = 2^n\right) = \frac{2^{n-1}!2^{2^{n-1}}}{2^n!}\cdot e^{-\frac{1}{2}} = 2{,}97\cdot 10^{-7}..$$

Для полного набора ячеек таблицы (исключая нулевой столбец и нулевую строку) имеем

$$2{,}97\cdot 10^{-7}\cdot(2^4-1)^2=0{,}000067,$$

т.е. вероятность получить вырожденную полубайтовую подстановку путём отбора получается менее семи десятитысячных.

Для байтовой подстановки соответственно получаем

$$\Pr\left(\Lambda_\pi(\Delta X,\Delta Y)=2^8\right)=9\cdot 10^{-254}$$

и тогда

$$9\cdot 10^{-254}\cdot(2^8-1)^2=5{,}85\cdot 10^{-249}.$$

В табл. 7 представлены результаты вычисления значений "хвостов" интегрального закона распределения для распределения вероятностей (2), т.е. считалась сумма $\sum_{k=k^*+1}^{2^{n-1}}\Pr(\Lambda_\pi(\Delta X,\Delta Y)=2k)$ как функции размера битового входа подстановки n и максимального половинного значения дифференциала k^*.

Таблица 7

n	k^*	$\sum_{k=k^*+1}^{2^{n-1}}\Pr(\Lambda_\pi(\Delta X,\Delta Y)=2k)$
4	3	0,00245 (0,55)
6	5	0,000015 (0,06)
8	6	0,000001 (0,065)
10	7	0,00000006 (0,062)
12	8	0,0000000034 (0,0057)

Видно, что даже в "нормированном" варианте доля сумм "хвостов" оказывается незначительной (а вырожденные подстановки попадают именно в хвосты). В скобках строчек табл. 7 приведены результаты умножения "хвостов" на общее число "ненулевых" ячеек таблицы $(2^n-1)^2$.

О подстановке под номером 6. В качестве эксперимента взяли хорошую подстановку шифра baby Rijndael [12] и с помощью двух транспозиций воспроизвели в ней переходы 2→4, 3→6 и 4→8 (последний переход уже был в исходной подстановке), имеющиеся в подстановке под номером 6. Полученная в результате этого подстановка повторила по свойствам свойства подстановки под номером 6 (пришла к асимптотическому значению максимума дифференциала, равному 24), т.е. стала вырожденной. В то же время одна транспозиция в подстановке под номером 6 в цикле, не содержащем отмеченные выше переходы, сделала подстановку невырожденной. Нам не удалось найти признаков, по которым можно делить подстановки на вырожденные и невырожденные, однако нам не удалось найти невырожденных подстановок среди наиболее вероятного множества подстановок, приближающихся по линейным и дифференциальным показателям к показателям случайной подстановки, определяемой законами распределения вероятностей (1) и (2).

Представленные примеры вырожденных подстановок ярко свидетельствуют, что S-блоки в шифрах играют важную роль. Существуют подстановки (вырожденного типа), с которыми построить хорошее криптографическое преобразование нельзя. С другой стороны, подстановки являются одним из важных элементов шифрующего преобразования. Они реализуют один из важных для шифра механизмов – механизм нелинейного перемешивания (перестановки) битов блоков данных, с помощью которого удаётся наиболее просто добиться эффекта хаотичности в преобразовании битов.

Выводы

К вырожденным S-блокам мы отнесли подстановочные конструкции с дифференциальными и линейными показателями (максимумами XOR таблиц и смещений таблиц линейных аппроксимаций) близкими к предельно возможным.

Результатами работы подтверждено, что получение вырожденных S-блоков при случайном порождении подстановок является маловероятным событием. Особенно это относится к байтовым S-блокам. Для этих S-блоков получение подстановок с максимумами XOR таблиц и смещений таблиц линейных аппроксимаций, близкими к предельно достижимым, является практически невозможным событием. Реальные наиболее вероятные значения максимумов, которые удаётся получить в экспериментах (34 для ЛАТ и 10-12 для ТР) оказываются далёкими от значений $2^{8-1}=128$ (для ЛАТ) и $2^{8}=256$ (для ТР). При этом с увеличением значений максимумов линейных и дифференциальных показателей вероятности отбора подстановок с такими значениями очень быстро уменьшаются. Таким образом, доля вырожденных подстановок в общем множестве подстановок симметрической группы оказывается незначительной.

Это значит, что положение, сформулированное в начале работы, состоящее в том, что все шифры независимо от используемых в них S-блоков после небольшого начального числа циклов шифрования становятся случайными подстановками, выполняется с высоким уровнем доверия. Порождение вырожденных S-блоков является маловероятным событием, и самое главное – они всегда могут быть обнаружены и исключены на основе результатов экспериментов.

Список литературы: 1. *Лисицкая, И.В.* Методология оценки стойкости блочных симметричных шифров / И.В. Лисицкая // Автоматизированные системы управления и приборы автоматики. – 2011. – № 163. – С. 123-133. 2. *Лисицкая, И.В.* Большие шифры – случайные подстановки / И.В. Лисицкая, А.А. Настенко // Радиотехника. – 2011. – Вып. 166. – С. 50-55.6. 3. *Лисицкая, И.В.* Дифференциальные свойства шифра FOX / И.В. Лисицкая, Д. С. Кайдалов // Прикладная радиоэлектроника. – 2011. – Т.10, №2. – С. 122-126. 4. *Markku-Juhani O. Saarinen* Cryptographic Analysis of All 4×4-Bit S-Boxes // IACR Cryptology ePrint Archive Vol. 2011 (2011), p. 218. 5. *Токарева, Н. Н.* Квадратичные аппроксимации специального вида для четырёхразрядных подстановок в S-блока // Прикладная дискретная математика. – 2008. – Т. 1, N 1. – С. 50-54. 6. *H. M. Heys.* A Tutorial on Linear and Differential Cryptanalysis, CRYPTOLOGIA, v 26, N 3, 2002, p 189-221. 7. *Лисицкая, И.В.* Об участии S-блоков в формировании максимальных значений линейных вероятностей блочных симметричных шифров / И.В. Лисицкая, В.В. Ковтун // Радиотехника. – 2011. – Вып. 166. – С. 17-25. 8. *Лисицкая, И.В.* Об участии S-блоков в формировании максимальных значений дифференциальных вероятностей блочных симметричных шифров / Лисицкая И.В., Казимиров А.В. // Proceedings International Conference SAIT 2011, Kyiv, Ukraine, May 23-28. – 2011. – С. 459. 9. *Долгов, В.И.* Свойства таблиц линейных аппроксимаций случайных подстановок / В.И. Долгов, И.В. Лисицкая, О.И. Олешко // Прикладная радиоэлектроника. – 2010. – Т. 9, № 3. – С. 334-340. 10. *Joan Daemen.* Probability distributions of Correlation and Differentials in Block Ciphers / Joan Daemen, Vincent Rijmen // April 13, 2006, pp. 1–38. 11. *Олейников, Р.В.* Дифференциальные свойства подстановок / Р.В. Олейников, О.И. Олешко, К.Е. Лисицкий, А.Д. Тевяшев // Прикладная радиоэлектроника. – 2010. – Т.9. – № 3. – С. 326-333. 12. *Долгов, В.И.* Исследование криптографических свойств нелинейных узлов замены уменьшенных версий некоторых шифров / В.И. Долгов, А.А. Кузнецов и др. // Прикладная радиоэлектроника. – 2009. – Т. 8, № 3. – С. 268-277.

Харьковский национальный
университет радиоэлектроники

Поступила в редколлегию 17.09.2012

УДК 681.3.06: 519.248.681

В.И. РУЖЕНЦЕВ, канд. техн. наук

ПРО ДОКАЗАТЕЛЬСТВО ОТСУТСТВИЯ ЭФФЕКТИВНЫХ БАЙТОВЫХ ДИФФЕРЕНЦИАЛЬНЫХ ХАРАКТЕРИСТИК ДЛЯ RIJNDAEL-ПОДОБНЫХ ШИФРОВ

Вопросы криптостойкости американского стандарта шифрования FIPS-197 [1] (AES или Rijndael [2]) вызывают интерес у многих криптографов. В работе [3] представлено обоснование стойкости этого алгоритма блочного симметричного шифрования к атаке усеченных (байтовых) дифференциалов. Однако позже были выявлены некоторые неточности в доказательстве одной из представленных в [3] теорем. В соответствии с этой теоремой определялось минимальное количество циклов, при котором для вариантов шифра rijndael со 128-, 192- и 256-битным блоками не существуют эффективные байтовые дифференциальные характеристики.

Цель настоящей работы – устранение неточностей и представление нового варианта доказательства теоремы, а также демонстрация результатов вычислительных экспериментов, которые подтверждают справедливость теоретических выводов, представленных в этой работе и в [3].

Методика реализации атаки усеченных дифференциалов предложена Л. Кнудсеном [4]. Отличие от обычной дифференциальной атаки заключается в том, что через циклы проводится не полная разность, а некоторая ее часть. В работе [4] показано, что такая методика эффективна в случаях, когда в шифре используется недостаточно хорошее рассеивание и прохождение разности через несколько циклов может рассматриваться независимо от значения разности в некоторой части блока.

Для байт-ориентированных шифров естественным считается изучение усеченных дифференциалов особого вида, для которых усечение заключается не в исключении из рассмотрения отдельных битов входной или выходной разности, а в рассмотрении активности S-блоков. Поскольку используемые в современных шифрах S-блоки (S-подстановки) чаще всего являются отображениями байт-в-байт, то дифференциалы, характеризующие активность S-блоков, часто называют байтовыми дифференциалами. В байтовых дифференциалах разность представляется в виде последовательности битов, каждый из которых отражает активность одного S-блока (байта) (1 – ненулевая разность – S-блок активный; 0 – S-блок неактивный). То есть, при рассмотрении байтовых дифференциалов (БД) в определенном смысле происходит объединение некоторого множества дифференциальных характеристик, поэтому распространено мнение [5], что оценка вероятностей БД для байт-ориентированных шифров позволяет получить более точную оценку стойкости к дифференциальному криптоанализу (ДК), чем при стандартном подходе, т.е. при рассмотрении максимальной вероятности отдельных дифференциальных характеристик.

В литературе существует описание атак усеченных (байтовых) дифференциалов на ослабленные варианты байт-ориентированных шифров SAFER и E2 [6,7].

Основываясь на [7 – 9], изложим основные понятия, связанные с усеченными (байтовыми) дифференциалами. Важными понятиями при изучении байтовых дифференциалов являются функция, или вес, Хэмминга (h-функция) и функция-характеристика (χ-функция). Вес Хэмминга битовой последовательности равен числу ненулевых битов в этой последовательности; χ-функция задает отображение $\{0,1\}n$ в $\{0,1\}$ следующим образом

$$\chi(a) = \begin{cases} 0, \text{если } a = 0, \\ 1, \text{если } a \neq 0, \end{cases}$$

где $a \in \{0,1\}^n$.

Если $x \in \{0,1\}nm$ то $\chi(x) = \chi(x0, x1, \ldots, xm) = (\chi(x0), \chi(x1), \ldots, \chi(xm))$.

Если обозначить разность на входе и выходе циклового преобразования, включающего в себя m n-битных S-блоков, как $\Delta x = (\Delta x_0, \Delta x_1, ..., \Delta x_m)$ и $\Delta y = (\Delta y_0, \Delta y_1, ..., \Delta y_m)$ соответственно, то векторы активизации байтов $\delta x, \delta y \in \{0,1\}^m$ определяются как

$$\delta x = (\delta x_0, \delta x_1, ..., \delta x_m), \text{ где } \delta x_i = \chi(\Delta x_i) \in \{0,1\},$$
$$\delta y = (\delta y_0, \delta y_1, ..., \delta y_m), \text{ где } \delta y_i = \chi(\Delta y_i) \in \{0,1\}.$$

Вероятность перехода $\delta x \to \delta y$ в ходе циклового преобразования F для любых $\delta x, \delta y \in \{0,1\}^m$ и $\Delta x, \Delta y \in \{0,1\}^{nm}$ определяется следующим образом

$$p^{(1)}_{бд}(\delta x, \delta y) = \frac{\sum_{\substack{\chi(\Delta x)=\delta x \\ \chi(\Delta y)=\delta y}} \Pr[F(x) \oplus F(x+\Delta x) = \Delta y]}{\#\{\Delta x \in \{0,1\}^{nm} \mid \chi(\Delta x) = \delta x\}}.$$

Совокупность значений входного и выходного векторов активизации для одного цикла преобразований называется *одноцикловой байтовой характеристикой*. Такая связь характеризуется вероятностью, которую будем обозначать как $p^{(1)}_{бд}(\delta x, \delta y)$, где δx и δy – входной и выходной вектора активизации. По аналогии с обычным дифференциальным криптоанализом «сшивку» нескольких одноцикловых байтовых характеристик (условие «сшивки»: входной вектор активизации каждой последующей одноцикловой байтовой характеристики равен выходному вектору активизации предыдущей) будем называть *многоцикловой байтовой характеристикой*. Вероятность такой характеристики вычисляется как произведение вероятностей всех входящих в нее одноцикловых характеристик. Байтовые характеристики, покрывающие одинаковое число циклов и имеющие одинаковые значения входных векторов активизации и одинаковые значения выходных векторов активизации, принадлежат одному и тому же *байтовому дифференциалу*. Вероятность байтового дифференциала есть сумма вероятностей всех входящих в него байтовых характеристик.

В работе [8] японские криптоаналитики определяют вероятность i-циклового усеченного дифференциала так:

$$P^{(i)}_{бд}(\beta'(0), \beta'(i)) = P(\chi(\Delta X(i)) = \beta'(i) \mid \chi(\Delta X(0)) = \beta'(0)),$$

где χ – функция-характеристика; $\beta' = \chi(\beta(i))$, а $\beta(i)$ – возможная дифференциальная разность на выходе i-го нелинейного уровня; $\Delta X(0)$ – входная разность; $\Delta X(i)$ – разность на выходе $(i+1)$-го цикла.

Как следует из [6, 7], наличие r-циклового эффективного байтового дифференциала позволяет организовать атаку на r-цикловый, а в некоторых случаях и на $(r+1)$-цикловый шифр.

Напомним также, что байтовая дифференциальная характеристика или байтовый дифференциал считаются *эффективными*, когда вероятность *рбд* или *Рбд* больше вероятности получения на выходе того же вектора активизации при произвольном (случайном) векторе активизации на входе (случайный входной вектор активизации предполагает равновероятность всех значений выходной разности):

$$Рбд > рсл \text{ или } рбд > рсл, \quad (1)$$

где $рсл \approx \left(2^{-8}\right)^u$, u – число неактивных байтов в выходной разности или число нулевых битов в выходном векторе активизации. Следует отметить, что для эффективного байтового дифференциала (эффективной байтовой характеристики) непременно будет выполняться и традиционное для обычных дифференциалов ограничение: $Рбд > 2{-}n$ ($рбд > 2{-}n$), где n – длина блока в битах.

Если удается найти эффективный дифференциал, то обычно на последнем цикле, где известны выходное значение разности (значения криптограмм) и входной вектор активизации (в соответствии с используемым байтовым дифференциалом). Эта информация позволяет получить информацию о подключе последнего цикла.

В основе выполнения оценки стойкости БСШ к атаке усеченных дифференциалов обычно лежит поиск эффективных байтовых дифференциалов, покрывающих достаточное для построения атаки число циклов и обладающих достаточно высокой вероятностью.

Представленную в работе [3] лемму 1 следует переформулировать. Вместо представленной в [3] формулировки:

«Для Rijndael-подобного шифра с k колонками в блоке нет эффективных байтовых дифференциальных характеристик с k или более активными колонками на входе преобразований MixColumns»,

справедливой следует считать

«Для Rijndael-подобного шифра с k колонками в блоке, эффективная байтовая дифференциальная характеристика не может содержать ни одного цикла, в котором были бы активными все k колонок на входе преобразования MixColumns».

Доказательство леммы [3] выполнено именно для новой формулировки.

Такая формулировка затрудняет возможность определения граничного количества циклов для существования эффективных БДХ, так как для 128-битного rijndael, например, можно сколько угодно циклов подряд держать лишь три активные колонки на входе MixColumns. Такая формулировка дает возможность доказать отсутствие эффективных БДХ для тех случаев, когда количество колонок в блоке 1 или 2. Для определения граничного количества циклов для большинства вариантов шифра нужно прибегнуть к вычислительным экспериментам по поиску эффективных БДХ с использованием, например, методики из [3], модифицированной для SPN-шифров. Для размера же блока 128 битов (4 32-битовые колонки в блоке) в этой работе представлен новый вариант доказательства отсутствия эффективных БДХ.

В работе [2] для шифра rijndael доказана лемма о числе активных колонок в двух последовательных циклах шифра. Согласно этой лемме, суммарное количество активных колонок на входе и выходе двух циклов преобразований составляет не менее 5.

Используя эту лемму, покажем, что для шифра FIPS-197 (rijndael со 128-битным блоком) справедлива следующая теорема.

Теорема. Для шифра FIPS-197 не существует эффективных байтовых дифференциальных характеристик (БДХ) для трех и более циклов.

Доказательство. Рассмотрим трехцикловую БДХ для такого шифра. Пусть на входе МС-преобразования 1-го, 2-го и 3-го цикла будет соответственно a, b и c активных колонок. ($a > 0$, $b > 0$ и $c > 0$). В первом цикле после операции МС в каждой активной колонке должно быть не более b активных байтов (в противном случае во втором цикле будет больше, чем b активных колонок на входе МС). Тогда, учитывая, что каждый пассивный байт на выходе преобразования МС уменьшает вероятность БДХ примерно в 28 раз, верхняя граница вероятности БДХ после первого цикла составит

$$2^{-(4-b)\cdot 8a}.$$

На основе аналогичных соображений после двух циклов вероятность БДХ составит не более чем

$$2^{-(4-b)\cdot 8a} \cdot 2^{-(4-c)\cdot 8b}.$$

При этом, если считать, что в третьем цикле после преобразования МС в каждой из c активных колонок будут активны все байты, то, используя выражение (1) для вычисления нижней границы вероятности эффективной БДХ, можно получить $pcл = 2^{-(4-c)\cdot 32}$.

Для того чтобы теорема была справедлива, должно выполняться неравенство

$$2^{-(4-b)\cdot 8a} \cdot 2^{-(4-c)\cdot 8b} \le 2^{-(4-c)\cdot 32}. \qquad (2)$$

Следует отметить, что каждый дополнительный пассивный байт на выходе операции MC третьего цикла будет добавлять множитель 2–8 в обе части этого неравенства.

Неравенство (2) равносильно неравенству

$$-(4-b)\cdot 8a-(4-c)\cdot 8b \le -(4-c)\cdot 32. \qquad (3)$$

В соответствии с леммой о числе активных колонок [2] можно записать $a+c \ge 5$. Поскольку $4 \ge b$, то $(4-b) \ge 0$, а следовательно $-(4-b)\cdot 8a \le 0$. Поэтому в выражение (3) можно подставить вместо a минимальное значение, т.е. сделать замену $a = 5-c$. Раскрыв после этого в (3) скобки и выполнив элементарные преобразования, приходим к равносильному неравенству $8b \le 32$, которое является справедливым, так как в соответствии с утверждением $b < 4$. Теорема доказана.

Справедливость доказанной теоремы подтвердилась в ходе вычислительных экспериментов по поиску эффективных БДХ с использованием метода Мораи [9] и метода, предложенного в работе [10]. Результаты поиска для шифров с полным набором преобразований во всех, включая последний, циклах представлены в таблице:

Размер блока шифра, биты	Число колонок в блоке	Максимальное число циклов	Максимальная вероятность БДХ, *рбд*	*рбд/рсл*
128	4	2	≈ 2–8	1,68e+07
192	6	3	≈ 2–16	62991
256	8	5	≈ 2–80	60101

Из полученных в ходе вычислительных экспериментов данных понятно, что для вариантов шифра rijndael с размером блока 192 и 256 битов может быть доказана аналогичная теорема (аналог теоремы 1 из [3]) об отсутствии эффективных БДХ для 4 и 6 циклов соответственно.

Дальнейший ход доказательства отсутствия эффективных байтовых дифференциалов, представленный в [3], остается прежним.

Основным итогом работы стало обоснование отсутствия эффективных БДХ для шифра rijndael-128 с числом циклов 3 и более. Используя представленные в [3] аргументы, можно утверждать, что шифр AES с еще одним дополнительным циклом (4 и более циклов) является защищенными от атаки усеченных (байтовых) дифференциалов.

Список литературы: 1. *National* Institute of Standards and Technology "Advanced encryption algorithm (AES) development" // FIPS 197, U.S. Department of Commerce, Nov. 2001. 2. *J. Daemen, V. Rijmen.* AES Proposal Rijndael, AES Round 1 Technical Evaluation CD-1: Documentation, National Institute of Standards and Technology, Aug 1998. See http://www.nist.gov/aes. 3. *Руженцев, В.И.* Доказуемая стойкость rijndael-подобных шифров к атаке усеченных дифференциалов // Радіоелектронні і комп'ютерні системи. – 2012. – №5. – С. 51-55. 4. *L. R. Knudsen.* Truncated and Higher Order Differentials. In B. Preneel, editor, Fast Software Encryption – Second International Workshop, Volume 1008 of Lecture Notes in Computer Science, pp. 196–211. Springer-Verlag, Berlin, Heidelberg, New York, 1995. 5. *K. Aoki, T. Ichikawa, M. Kanda, M. Matsui, S. Moriai, J. Nakajima, T. Tokita.* Camellia: A 128-Bit Block Cipher Suitable for Multiple Platforms, Selected Areas in Cryptography – 7th Annual International Workshop, SAC2000, Lecture Notes in Computer Scienca 2012, pp. 39-56, Springer-Verlag, Berlin, 2001. 6. *L. R. Knudsen, T. A. Berson.* Truncated differentials of SAFER, In Fast Software Encryption – Third International Workshop, FSE'96, Volume 1039 of Lecture Notes in Computer Science, Berlin, Heidelberg, New York, Springer-Verlag, 1996. 7. *M. Matsui, T. Tokita.* Cryptanalysis of reduced version of the block cipher E2, in pre-proceedings of Fast Software Encryption'99, pp. 70-79, 1999. 8. *M. Sugita, K. Kobara.* Relationships among differential, truncated differential, impossible differential cryptanalyses against word-oriented block cipher like Rijndael, E2 // National Institute of Standards and Technology, http://www.nist.gov/aes. 9. *S. Moriai, M. Sugita, K. Aoki.* Security of E2 against Truncated Differential Cryptanalysis. In H. Heys and C. Adams, editors, Selected Areas in Cryptography — 6th Annual International Workshop, SAC'99, Volume 1758 of Lecture Notes in Computer Science, pp. 106–117, Berlin, Heidelberg, New York, Springer-Verlag, 2000. 10. *Руженцев, В.И.* О методах оценки стойкости к атаке усеченных дифференциалов // Радиоэлектроника и информатика. – 2003. – №4. – С. 130-133.

Харьковский национальный университет радиоэлектроники

Поступила в редколлегию 09.09.20012

УДК 681.3.06

А.А. КУЗНЕЦОВ, д-р техн. наук, А.В. КОВАЛЕНКО, канд. техн. наук, С.А. ИСАЕВ

ФОРМИРОВАНИЕ НЕЛИНЕЙНЫХ УЗЛОВ ЗАМЕН С ИСПОЛЬЗОВАНИЕМ НЕДВОИЧНЫХ КРИПТОГРАФИЧЕСКИХ ФУНКЦИЙ

Введение

Регулярные нелинейные криптографические функции (узлы замен) симметричных шифров реализуют отображение n-битных блоков входных данных в m-битные выходные блоки: $F: GF^n(2) \to GF^m(2)$. Традиционный подход к описанию, оцениванию и разработке методов синтеза регулярных нелинейных узлов замен состоит в представлении функции F с помощью ее координатных функций, которые задаются в терминах булевой алгебры [1]. В то же время, как показано в [2, 3], построение нелинейных узлов замен с высокими показателями стойкости через итеративное формирование компонентных булевых функций является непрактичным уже при $n = 6$ и вычислительно недостижимым для $n > 6$. Это предполагает обоснование новых подходов к описанию криптографических узлов замен симметричных шифров, исследование математического аппарата оценивания основных показателей стойкости и построение вычислительно эффективных алгоритмов синтеза.

Представление S-блоков через компонентные булевы функции

Введем основные понятия и определения математического аппарата булевой алгебры, используемые в дальнейшем при описании нелинейных узлов замен через компонентные булевы функции и оценке их криптографических свойств.

Булевой функцией $f(x_1,...,x_n)$ от n переменных является функция, осуществляющая отображение из поля $GF(2^n)$ всех двоичных векторов $x = (x_1,...,x_n)$ длины n в поле $GF(2)$ [4]. Обычно булевы функции представляются в алгебраической нормальной форме (АНФ), т.е. рассматриваются как сумма произведений составляющих координат:

$$f(x_1,...,x_n) = \lambda_0 + \lambda_1 x_1 + ... + \lambda_n x_n + +\lambda_{12} x_1 x_2 + \lambda_{13} x_1 x_3 + ... + \lambda_{12...n} x_1 x_2 ... x_n, \qquad (1)$$

где $\lambda_0, \lambda_1, ..., \lambda_{12}, ..., \lambda_{12...n}$ – уникальные двоичные константы, а суммирование и умножение производится в двоичном поле $GF(2)$.

Поле $GF(2^n)$ состоит из 2^n векторов $\alpha_i = (\alpha_1^i, \alpha_2^i, ..., \alpha_n^i)$, $\alpha_j^i \in GF(2)$: $\alpha_0 = (0,...,0,0)$, $\alpha_1 = (0,...,0,1)$, …, $\alpha_{2^n-1} = (1,...,1,1)$, $\alpha_i \in V_n$, где V_n – векторное пространство над $GF(2)$.

Таблицей истинности функции f называется (0,1)-последовательность, определенная как [5]:

$$\left(f(x) \middle| x \in GF^n(2)\right) = \left(f(\alpha_0), f(\alpha_1), ..., f(\alpha_{2^n-1})\right).$$

Последовательностью функции f, обозначаемой $\tilde{f}$, называется (1,-1)-последовательность, определенная как [5]:

$$\left((-1)^{f(x)} \middle| x \in GF^n(2)\right) = \left((-1)^{f(\alpha_0)}, (-1)^{f(\alpha_1)}, ..., (-1)^{f(\alpha_{2^n-1})}\right).$$

Рассмотрим криптографические свойства функций, реализующих отображения из $GF^n(2)$ в $GF^m(2)$, где $1 \le m \le n$. Пусть M_n^m есть множество таких функций, а B_n есть множество булевых функций от n переменных, то есть функций, реализующих отображения

из $GF^n(2)$ в $GF(2)$. Тогда любую функцию $F \in M_n^m$ можно рассматривать как состоящую из m булевых функций от n переменных, т.е. m-выходных координатных функций из B_n.

В более общем представлении, компонентная функция $F \in M_n^m$ является ненулевой линейной комбинацией ее координатных функций из B_n.

Таким образом, функцию $F: GF^n(2) \to GF^m(2)$ запишем через множество $F = (f_1(x_1,...,x_n),...,f_m(x_1,...,x_n))$, где $f_i(x_1,...,x_n) \in B_n$.

Алгебраическая степень f [5], обозначаемая $\deg(f)$, определяется как максимальная степень многочлена представленного в АНФ.

Важные свойства булевых функций изучаются с использованием преобразования Уолша – Адамара.

Преобразование Уолша – Адамара функции $f(x_1,...,x_n) \in B_n$ есть вещественная функция $\overline{F}(w)$ [5]:

$$\overline{F}(w) = \sum_{x \in GF^n(2)} (-1)^{f(x)+w \cdot x}, \tag{2}$$

где скалярное произведение векторов x и w определяется как $x \cdot w = x_1 w_1 + ... + x_n w_n$.

Булева функция f сбалансирована, если вероятности событий $f(x) = 1$ и $f(x) = 0$ равны. Используя преобразование Уолша – Адамара, условие сбалансированности функции f запишем в виде $\overline{F}(0) = 0$.

Расстояние по Хеммингу между двумя функциями f и g из B_n определяется как:

$$d_H(f,g) = card\left\{x \mid f(x) \neq g(x), x \in GF^n(2)\right\}. \tag{3}$$

Нелинейность $NL(f)$ функции $f(x_1,...,x_n) \in B_n$ определяется как [5]:

$$NL(f) = \min_{g \in A_n} d_H(f,g), \tag{4}$$

где A_n – множество всех аффинных функций от n переменных,

$$A_n = \left\{ a_0 + \sum_{i=1}^{n} a_i x_i \mid a_i \in GF(2), 0 \le i \le n \right\}. \tag{5}$$

С использованием преобразования Уолша – Адамара нелинейность функции f может быть получена следующим образом:

$$NL(f) = 2^{n-1} - \frac{1}{2} \max_{w \in GF^n(2)} \left|\overline{F}(w)\right|. \tag{6}$$

Взаимосвязь показателя нелинейности $NL(f)$ функции $f(x_1,...,x_n) \in B_n$ с преобразованием Уолша – Адамара и вывод формулы (6) легко понять, представив выражение (2) в виде матричного умножения последовательности функции $\left((-1)^{f(x)} \middle| x \in GF^n(2)\right)$, на матрицу Уолша – Адамара H_{2^n} порядка 2^n:

$$\left(\overline{F}(w) \middle| w \in GF^n(2)\right) = \left(\sum_{x \in F_2^n} (-1)^{f(x)+w \cdot x} \middle| w \in GF^n(2)\right) = \left((-1)^{f(x)} \middle| x \in GF^n(2)\right) \cdot H_{2^n}$$

(последовательность функции в данном выражении и далее по тексту представляется в виде вектора-строки, образованной элементами этой последовательности).

Итеративное правило построения матрицы H_{2^n} задается выражением

$$H_1 = \left|1\right|,\ H_{2^i} = \begin{vmatrix} H_{2^{i-1}} & H_{2^{i-1}} \\ H_{2^{i-1}} & -H_{2^{i-1}} \end{vmatrix},\ i \in N.$$

Каждая строка матрицы Уолша – Адамара соответствует последовательности некоторой аффинной функции $g_i(x_1,...,x_n)$ из A_n с $a_0 = 0$ в общем представлении (5). Строго говоря, полное множество последовательностей всех аффинных функций с $a_0 = 0$ упорядочены по строкам (столбцам) матрицы Уолша – Адамара естественным образом:

$$A'_n = \left\{ \begin{array}{c} g_1(x_1,...,x_n) = 0 \\ g_2(x_1,...,x_n) = x_1 \\ g_3(x_1,...,x_n) = x_2 \\ ... \\ g_{2^n}(x_1,...,x_n) = x_1 + x_2 + ... + x_n \end{array} \right\},$$

где $g_i(x_1,...,x_n)$ – i-я аффинная функция, из упорядоченного подмножества аффинных функций A'_n с $a_0 = 0$ в (5).

Другими словами, последовательность $\left((-1)^{g_i(x)} \middle| x \in GF^n(2)\right)$ i-й аффинной функции из A'_n соответствует i-й строке матрицы Уолша – Адамара и наоборот.

Тогда выполняется равенство

$$\left((-1)^{f(x)} \middle| x \in GF(2)^n\right) \cdot H_{2^n} = \left((-1)^{f(x)} \middle| x \in GF^n(2)\right) \cdot \left(\begin{array}{c} \left((-1)^{g_1(x)} \middle| x \in GF^n(2)\right) \\ \left((-1)^{g_2(x)} \middle| x \in GF^n(2)\right) \\ ... \\ \left((-1)^{g_{2^n}(x)} \middle| x \in GF^n(2)\right) \end{array} \right)^T =$$

$$= \left(\sum_{x \in GF^n(2)} (-1)^{f(x)+g_i(x)} \middle| x \in GF^n(2),\ g_i(x) \in A'_n \right) = \left(\overline{F}(w) \middle| w \in GF^n(2) \right).$$

Например, для $n = 2$ имеем матрицу Уолша – Адамара H_4:

$$H_4 = \begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{vmatrix},$$

причем

$$\left((-1)^{g_1(x)=0} \middle| x \in GF^2(2)\right) = (1,1,1,1);\ \left((-1)^{g_2(x)=x_1} \middle| x \in GF^2(2)\right) = (1,-1,1,-1);$$

$$\left((-1)^{g_3(x)=x_2}\, x \in GF^2(2)\right) = (1,1,-1,-1);\ \left((-1)^{g_4(x)=x_1+x_2}\,\middle|\, x \in GF^2(2)\right) = (1,1,-1,-1)$$

и матричное произведение $\left((-1)^{f(x)}\middle| x \in GF^2(2)\right)\cdot H_4$ соответствует вычислению вектора значений функции $\overline{F}(w)$ для всех $w \in GF^2(2)$.

Выражение для расчета значений коэффициентов преобразования Уолша – Адамара запишем в виде

$$\overline{F}(w) = \sum_{x \in GF^n(2)} (-1)^{f(x)+g_w(x)}\middle| x \in GF^n(2).$$

Максимальное значение коэффициентов преобразования Уолша – Адамара $\max\limits_{w \in GF^n(2)} \overline{F}(w)$ булевой функции $f(x)$ соответствует максимальному коэффициенту корреляции (похожести) последовательности этой функции и последовательностей всех аффинных функций из множества A'_n:

$$\max_{w \in GF^n(2)} \overline{F}(w) = \max_{w \in GF^n(2)} \left(\sum_{x \in GF^n(2)} (-1)^{f(x)+g_w(x)}\middle| x \in GF^n(2) \right) =$$

$$= \max_{w \in GF^n(2)} \begin{pmatrix} card\left\{x \mid f(x) = g_w(x), x \in GF^n(2)\right\} - \\ -card\left\{x \mid f(x) \neq g_w(x), x \in GF^n(2)\right\} \end{pmatrix}.$$

Последовательности аффинных функций с $a_0 = 1$ в (5) соответствуют инверсии (умножению на «-1») последовательностей функций из A'_n, следовательно, максимум модуля коэффициентов преобразования Уолша – Адамара $\max\limits_{w \in GF^n(2)} \left|\overline{F}(w)\right|$ булевой функции $f(x)$ будет соответствовать максимальному коэффициенту корреляции последовательности этой функции и последовательностей всех аффинных функций из множества A_n.

По определению нелинейности из (4) имеем:

$$NL(f) = \min_{g \in A_n} d_H(f,g) = \min_{g \in A_n} card\left\{x \mid f(x) \neq g(x), x \in GF^n(2)\right\}.$$

Поскольку $card\left\{x \mid f(x) = g(x), x \in GF^n(2)\right\} + card\left\{x \mid f(x) \neq g(x), x \in GF^n(2)\right\} = 2^n$ справедливо равенство

$$\max_{w \in GF^n(2)} \left|\overline{F}(w)\right| = \max_{w \in GF^n(2)} \left| \sum_{x \in GF^n(2)} (-1)^{f(x)+g_w(x)}\middle| x \in GF^n(2)\right| =$$

$$= \max_{w \in GF^n(2)} \left| \begin{matrix} card\left\{x \mid f(x) = g_w(x), x \in GF^n(2)\right\} - \\ -card\left\{x \mid f(x) \neq g_w(x), x \in GF^n(2)\right\} \end{matrix} \right| = \max_{w \in GF^n(2)} \left| 2^n - 2card\left\{x \mid f(x) \neq g(x), x \in GF^n(2)\right.\right.$$

$$= 2^n - 2\min_{g \in A_n} card\left\{x \mid f(x) \neq g(x), x \in GF^n(2)\right\} = 2^n - 2\min_{g \in A_n} d_H(f,g) = 2^n - 2N(f),$$

откуда имеем

$$NL(f) = \frac{2^n - \max\limits_{w \in GF^n(2)} \left|\overline{F}(w)\right|}{2} = 2^{n-1} - \max\limits_{w \in GF^n(2)} \left|\overline{F}(w)\right|.$$

Автокорреляционная функция, обозначаемая $r_{\hat{f}}(\alpha)$, вычисляется по формуле [6]:

$$r_{\hat{f}}(\alpha) = \sum_{x \in GF^n(2)} \hat{f}(x)\hat{f}(x \oplus \alpha),$$

где $\alpha \in GF^n(2)$ и $r_{\hat{f}}(0) = 2^n$. Автокорреляционная функция является вектором, содержащим 2^n действительных значений в диапазоне $[(-2)^n, 2^n]$.

Автокорреляция AC функции f является максимальным абсолютным значением автокорреляционной функции [6]:

$$AC = \max_{\alpha \in GF^n(2), \alpha \neq 0} \left|r(\alpha)\right|.$$

Таким образом, математический аппарат булевых функций является удобным инструментом для описания регулярных нелинейных узлов замен, а использование преобразования Уолша – Адамара дает адекватный механизм оценки основных криптографических показателей стойкости, в частности нелинейности компонентных булевых функций.

В то же время, использование рассмотренного математического аппарата для синтеза регулярных узлов замен через итеративное формирование компонентных булевых функций является непрактичным уже при $n = 6$ и вычислительно недостижимым для $n > 6$ [2, 3]. Перспективным направлением в этом смысле является использование недвоичных криптографических функций, описывающих отображение n-битных блоков входных данных в m-битные выходные блоки в нелинейном узле замен в виде функций отображения $F : GF(2^n) \to GF(2^m)$.

Представление S-блоков через недвоичные криптографические функции

Введем основные понятия и определения предлагаемого математического аппарата для описания нелинейных узлов замен через недвоичные функции.

Недвоичной (над полем $GF(2^{n_1})$) *функцией* $F(X_1,...,X_{n_2})$ от n_2 переменных является функция, осуществляющая отображение из поля $GF((2^{n_1})^{n_2})$ всех векторов $X = (X_1,...,X_{n_2})$ длины n_2 с элементами из $GF(2^{n_1})$ в поле $GF(2^{n_1})$. Как и рассмотренные выше булевы функции, каждая недвоичная функция $F(X_1,...,X_{n_2})$ может быть представлена в АНФ, т.е. как сумма произведений составляющих координат:

$$F(X_1,...,X_{n_2}) = \Lambda_0 + \Lambda_1 X_1 + ... + \Lambda_n X_n + \Lambda_{12} X_1 X_2 + \Lambda_{13} X_1 X_3 + ... + \Lambda_{12...n_2} X_1 X_2 ... X_n, \quad (7)$$

где $\Lambda_0, \Lambda_1, ..., \Lambda_{12}, ..., \Lambda_{12...n_2}$ – уникальные константы из $GF(2^{n_1})$, суммирование и умножение также производится в поле $GF(2^{n_1})$.

Поле $GF((2^{n_1})^{n_2})$ состоит из $2^{n_1 n_2}$ векторов $A_i = (A_1^i, A_2^i, ..., A_{n_2}^i)$, $A_j^i \in GF(2^{n_1})$:

$$A_0 = (0,...,0,0),\ A_1 = (0,...,0,1),\ \ldots,\ A_{2^{n_1}-1} = (0,...,0,2^{n_1}-1),\ A_{2^{n_1}} = (0,...,1,0),$$

$$A_{2^{n_1}+1} = (0,...,1,1),\ \ldots,\ A_{(2^{n_1})^{n_2}-1} = (2^{n_1}-1,...,2^{n_1}-1,2^{n_1}-1),\ \alpha_i \in V_n,$$

где V_{n_2} – векторное пространство над $GF(2^{n_1})$.

Поле $GF((2^{n_1})^{n_2})$ изоморфно полю $GF(2^n)$, $n = n_1 n_2$, т.е. имеем взаимно-однозначное функциональное соответствие множества векторов $A_i = (A_1^i, A_2^i, ..., A_{n_2}^i) \in V_{n_2}$ с элементами из $GF(2^{n_1})$ и двоичных векторов $\alpha_i = (\alpha_1^i, \alpha_2^i, ..., \alpha_n^i) \in V_n$.

Таблицей истинности недвоичной (над полем $GF(2^{n_1})$) функции F называется последовательность с элементами из $GF(2^{n_1})$, определенная как

$$\left(F(X)\middle| X \in GF^{n_2}(2^{n_1})\right) = \left(F(A_0), F(A_1), ..., F(A_{2^{n_1 n_2}-1})\right).$$

Последовательностью недвоичной (над полем $GF(2^{n_1})$) функции F называется последовательность из $2^{n_1 n_2}$ (1,-1)-кортежей длины $2^{n_1}-1$ каждый, определенная как

$$\left((-1)^{w\cdot F(X)}\middle| w \in GF(2^{n_1}), w \neq 0, X \in GF^{n_2}(2^{n_1})\right) =$$

$$= \left(\begin{array}{l} \left((-1)^{w_1\cdot F(A_0)}, (-1)^{w_2\cdot F(A_0)}, ..., (-1)^{w_{2^{n_1}-1}\cdot F(A_0)}\right), \\ \left((-1)^{w_1\cdot F(A_1)}, (-1)^{w_2\cdot F(A_1)}, ..., (-1)^{w_{2^{n_1}-1}\cdot F(A_1)}\right), ..., \\ \left((-1)^{w_1\cdot F(A_{2^{n_1 n_2}})}, (-1)^{w_2\cdot F(A_{2^{n_1 n_2}})}, ..., \right. \\ \left. (-1)^{w_{2^{n_1}-1}\cdot F(A_{2^{n_1 n_2}})}\right) \end{array}\right),$$

где w – накладываемая маска и $w \cdot F(X)$ – скалярное произведение векторных представлений чисел w и $F(X) \in GF(2^{n_1})$ (т.е. числа представлены в виде двоичных последовательностей, с элементами из $GF^{n_1}(2)$).

Например, пусть $n_1 = 2$, $n_2 = 1$ и недвоичная (над $GF(2^2)$) функция задана в АНФ следующим образом:

$$F(X) = 3 + X + 2X^2,$$

где коэффициенты многочлена принадлежат полю $GF(2^2)$: $0 = (0,0)$, $1 = (1,0)$, $2 = (0,1)$, $3 = (1,1)$.

Входными элементами такой функции являются однокоординатные вектора (скаляры) с элементами из $GF(2^2)$: $A_0 = (0)$, $A_1 = (1)$, $A_2 = (2)$, $A_3 = (3)$.

Таблицей истинности функции $F(X) = 3 + X + 2X^2$ является последовательность с элементами из $GF(2^2)$:

$$\left(F(X)\middle| X \in GF^1(2^2)\right) = \left(F(\mathrm{A}_0), F(\mathrm{A}_1), F(\mathrm{A}_2), F(\mathrm{A}_3)\right) = (3, 0, 0, 3).$$

Значения маски w и выхода функции $F(X)$ принадлежат полю $GF(2^2)$: $0 = (0,0)$, $1 = (1,0)$, $2 = (0,1)$, $3 = (1,1)$.

Последовательностью функции $F(X) = 3 + X + 2X^2$ является последовательность из $2^{n_1 n_2} = 4$ (1,-1)-кортежей длины $2^{n_1} - 1 = 3$ каждый:

$$\left(\left((-1)^{w_1 \cdot F(X)}, (-1)^{w_2 \cdot F(X)}, (-1)^{w_3 \cdot F(X)}\right)\middle| w \in GF(2^{n_1}), w \neq 0, X \in GF^1(2^2)\right) = \left(\begin{array}{l}\left((-1)^{w_1 \cdot F(A_0)}, (-1)^{w_2 \cdot F(A_0)}, (-1)^{w_3 \cdot F(A_0)}\right),\\ \left((-1)^{w_1 \cdot F(A_1)}, (-1)^{w_2 \cdot F(A_1)}, (-1)^{w_3 \cdot F(A_1)}\right),\\ \left((-1)^{w_1 \cdot F(A_2)}, (-1)^{w_2 \cdot F(A_2)}, (-1)^{w_3 \cdot F(A_2)}\right),\\ \left((-1)^{w_1 \cdot F(A_3)}, (-1)^{w_2 \cdot F(A_3)}, (-1)^{w_3 \cdot F(A_3)}\right)\end{array}\right) =$$

$$= \left(\begin{array}{l}\left((-1)^{(1,0)\cdot(1,1)}, (-1)^{(0,1)\cdot(1,1)}, (-1)^{(1,1)\cdot(1,1)}\right),\\ \left((-1)^{(1,0)\cdot(0,0)}, (-1)^{(0,1)\cdot(0,0)}, (-1)^{(1,1)\cdot(0,0)}\right),\\ \left((-1)^{(1,0)\cdot(0,0)}, (-1)^{(0,1)\cdot(0,0)}, (-1)^{(1,1)\cdot(0,0)}\right),\\ \left((-1)^{(1,0)\cdot(1,1)}, (-1)^{(0,1)\cdot(1,1)}, (-1)^{(1,1)\cdot(1,1)}\right)\end{array}\right) = \left(\begin{array}{l}\left((-1)^1, (-1)^1, (-1)^0\right), \left((-1)^0, (-1)^0, (-1)^0\right),\\ \left((-1)^0, (-1)^0, (-1)^0\right), \left((-1)^1, (-1)^1, (-1)^0\right)\end{array}\right) =$$

$$= \left((-1,-1,1), (1,1,1), (1,1,1), (-1,-1,1)\right).$$

Рассмотрим криптографические свойства функций F', реализующих отображения из $GF^{n_2}(2^{n_1})$ в $GF^m(2^{n_1})$, где $1 \le m \le n_2$. Пусть $M_{n_2}^m$ есть множество таких функций F', а B_{n_2} есть множество недвоичных функций $F(X_1, \ldots, X_{n_2})$ от n_2 переменных, то есть функций, реализующих отображения из $GF^{n_2}(2^{n_1})$ в $GF(2^{n_1})$. Тогда любую функцию из $M_{n_2}^m$ можно рассматривать как состоящую из m недвоичных функций $F(X_1, \ldots, X_{n_2})$ от n_2 переменных, т.е. m-выходных координатных функции из B_{n_2}. В более общем представлении, компонентная функция из $M_{n_2}^m$ является ненулевой линейной комбинацией ее координатных недвоичных функций из B_{n_2}.

Таким образом, функцию отображения $F' : GF^{n_2}(2^{n_1}) \to GF^m(2^{n_1})$, реализующую нелинейный узел замен, запишем через множество

$$F' = (F_1(X_1, \ldots, X_{n_2}), \ldots, F_m(X_1, \ldots, X_{n_2})),$$

где $F_i(X_1, \ldots, X_{n_2}) \in B_{n_2}$.

В данной работе ограничимся рассмотрением функций с $m = 1$, т.е. будем рассматривать только функции $F' = F_i(X_1, \ldots, X_{n_2})$, реализующие отображения из $GF^{n_2}(2^{n_1})$ в $GF(2^{n_1})$.

Введенная формализация функционального отображения $F: GF^{n_2}(2^{n_1}) \to GF(2^{n_1})$ является естественным обобщением рассмотренного выше подхода к представлению регулярных узлов замен в виде совокупности компонентных булевых функций.

Действительно, используя традиционный подход к описанию функционального отображения n-битных блоков входных данных в m-битные выходные блоки, функцию $F: GF^{n}(2) \to GF^{m}(2)$, где $n = n_1 n_2$, $m = n_1$ можно представить в виде кортежа $F = \{f_1(x_1,...,x_n), f_2(x_1,...,x_n),..., f_m(x_1,...,x_n)\}$ из $m = n_1$ булевых функций от $n = n_1 n_2$ булевых переменных каждая.

Для недвоичной функции из предыдущего примера имеем соответствие

$$F(X) = 3 + X + 2X^2 \equiv \begin{Bmatrix} f_1(x_1, x_2) = 1 + x_1 + x_2 \\ f_2(x_1, x_2) = 1 + x_1 + x_2 \end{Bmatrix},$$

где знак тождества означает тождественность правила отображения n = 2-битных блоков входных данных в m = 2-битные выходные блоки.

Важные свойства булевых функций изучаются с использованием преобразования Уолша – Адамара. По аналогии с преобразованием Уолша – Адамара введем спектральное преобразование недвоичных функций следующим образом.

Поскольку нелинейность булевой функции определяется как минимальное расстояние по Хэммингу от рассматриваемой функции ко множеству всех аффинных булевых функций, для недвоичного случая нам необходимо определить множество аффинных недвоичных функций.

Алгебраическая степень F, обозначаемая $\deg(F)$, определяется как максимальная степень многочлена представленного в АНФ.

По аналогии с классическим подходом назовем аффинной функцией недвоичную функция, чья алгебраическая степень $\deg(F) \le 1$. Соответственно, недвоичные функции, имеющие алгебраическую степень $\deg(F) > 1$, назовем нелинейными недвоичными функциями.

Спектральным преобразованием недвоичной функции $F(X_1,..., X_{n_2}) \in B_{n_2}$ есть вещественная функция $\overline{F}(W)$:

$$\overline{F}(W) = \sum_{X \in GF^{n_2}(2^{n_1})} \sum_{i=1}^{n_1} (-1)^{(F(X))_i + (G_W(x))_i}, \qquad (8)$$

где под $G_W(x)$ – понимается W-я недвоичная аффинная функция от n_2 переменных из множества A_n:

$$A_n \quad a_0 \quad \sum_{i\ 1}^{n_2} a_i X_i \mid a_i \quad GF(2^{n_1}), 0 \quad i \quad n \ . \qquad (9)$$

Так же, как и вектор w в случае булевого описания определяет вид линейных двоичных функций $g_w(x)$, в случае недвоичного описания вектор W задает вид недвоичных аффинных функций $G_W(x)$.

Вычислительный метод синтеза регулярных нелинейгых узлов замен

Анализ открытой литературы показал, что на сегодняшний день существуют вычислительные методы синтеза регулярных нелинейных узлов замен, среди которых [2, 8 - 12]: побитовые методы (bit-by-bit methods), методы случайной генерации с фильтрацией (random generation), метод градиентного подъема (hill climbing), генетические алгоритмы (genetic

algorithms), метод имитации отжига (simulated annealing), метод дифференциальной эволюции (differential evolution), метод оптимизации роем частиц (particle swarm optimization) и др. На наш взгляд, наиболее эффективным методом синтеза регулярных S-блоков является метод имитации отжига SA. Так, например, на основе проведенных в работах [10, 13] сравнений показано, что использование метода имитации отжига позволяет реализовать вычислительный поиск криптографических функций с лучшими на сегодняшний день показателями. Описание и алгоритм метода приведены ниже [10].

Поиск начинается с некоторого начального состояния $S = S_0$. Параметр T – некий контрольный параметр, известный как температура. T инициализируется высокой температурой T_0 и постепенно снижается. При каждом значении температуры выполняется определенное число *MIL* (Moves in Inner Loop, шагов во внутреннем цикле) шагов к новым состояниям. Состояние кандидата Y выбирается случайным образом из соседей $N(S)$ текущего состояния. Вычисляется изменение значения функции *cost*, $\delta = cost(Y) - cost(S)$. Если значение $cost(S)$ улучшается (т.е. $\delta < 0$ для задачи минимизации), тогда выполняется шаг относительно этого состояния ($S = Y$); в противном случае – он выполняется с некоторой вероятностью. Чем хуже шаг, тем меньше вероятность того, что он будет принят; чем ниже температура T, тем менее вероятно, что ухудшающий шаг будет принят. Вероятностное принятие решения определяется генерацией случайного числа U в интервале (0…1) и выполнением указанного ниже сравнения.

Вначале температура высокая и принимается почти каждый шаг. Это сделано для того, чтобы поиск носил не локальный, а глобальный характер. По мере того как температура уменьшается, становится все более трудно принимать ухудшающие шаги. В конце концов, допускаются только улучшающие шаги и процесс застывает. Алгоритм прерывается, когда встречается критерий остановки. Общий критерий остановки (который и был применен в нашей работе) – остановка поиска при достижении заданного числа *MaxIL* внутренних циклов, либо когда было выполнено некоторое максимальное число *MUL* последовательных непродуктивных внутренних циклов (т.е. без единого принятого шага). При этом лучшее достигнутое состояние сохраняется, поскольку поиск может выйти из него и впоследствии не найти состояние с подобными показателями. В конце каждого внутреннего цикла температура понижается. В нашей работе использовалось геометрическое охлаждение – умножение на константу охлаждения α в интервале (0…1).

Соседей функции f можно определить следующим образом. Функция g находится по соседству с функцией f, если:

$$\exists x, y \in Z_2^n : \hat{f}(x) \neq \hat{f}(y), \hat{g}(x) = \hat{f}(y), \hat{g}(y) = \hat{f}(x), \forall z \in Z_2^n \setminus \{x, y\} : \hat{g}(z) = \hat{f}(x).$$

Алгоритм имитации отжига *SA*

```
S = S0;
T = T0;
repeat {
  for (int i = 0; i < MIL; i++)
  {
        выбрать Y ∈ N(S);
        δ = cost(Y) – cost(S);
        if (δ < 0) then S = Y;
        else сгенерировать U = U(0,1);
        if (U < exp(-δ/T)) then S = Y;
  }
  T = T x α;
}
until (критерий остановки не достигнут).
```

Поиск начинался со сбалансированной, но при этом случайной функции. Один шаг алгоритма меняет местами два отличных элемента таблицы истинности функции, сохраняя ее сбалансированность.

Рассмотрим процедуры формирования функций стоимости cost (ценовых функций), используемые для синтеза S-блоков через спектральные характеристики булевых функций, введем соответствующие функции стоимости для синтеза S-блоков через спектры недвоичных криптографических функций.

Пусть функция $F(x): GF^n(2) \to GF^m(2)$ задает S-блок размерности *nxm*. Пусть для $\beta \in GF^m(2)$, $F_\beta(x) = \beta_1 f_1(x) \oplus ... \oplus \beta_m f_m(x)$ – линейная комбинация *m* выходов S-блока *F*. Тогда $\hat{F}_\beta(\omega), r_\beta(s)$ – значения преобразования Уолша – Адамара и значения автокорреляции для каждой булевой функции f_β.

Поскольку нелинейность булевой функции $NL(f) = \frac{1}{2}(2^n - \max_\omega |\hat{F}(\omega)|)$, то задача повышения нелинейности может быть представлена как задача минимизации абсолютного максимального значения коэффициента Уолша – Адамара. Изначально в задачах синтеза S-блоков по критерию высокой нелинейности для метода имитации отжига использовалась следующая функция стоимости [10]:

$$cost(f) = WHT_{\max}(f) = \max_\omega |\hat{F}(\omega)|.$$

Поскольку задача понижения автокорреляции представляется как задача минимизации максимального значения автокорреляционной функции, то *cost* функция в дальнейших исследованиях приняла вид [10]:

$$cost(f) = AC(f) = \max_{s \neq 0} |\sum_x \hat{f}(x)\hat{f}(x \oplus s)| = \max_{s \neq 0} |\hat{r}(s)|.$$

Обычно в многокритериальных задачах применяется следующий подход: вычисляется сумма отдельных *cost* функций (по различным критериям), умноженных на весовые коэффициенты. Тогда *cost* функция в задаче синтеза S-блока с высокой нелинейностью и низкой автокорреляцией принимает вид [10]:

$$\cos t(f) = \alpha \cdot WHT_{\max}(f) + \beta \cdot AC(f).$$

Далее были разработаны улучшенные функции, которые основывались на следующем положении.

Известно, что равенство Парсеваля

$$\sum_w (\hat{F}(w))^2 = 2^{2n}$$

ограничивает $WHT_{\max}(f) = \max_w \left|\hat{F}(w)\right|$ значением равным как минимум $2^{n/2}$. Данная граница достигается тогда, когда выполняется равенство $\left|\hat{F}(w)\right| = 2^{\frac{n}{2}}$ для каждого ω. Когда значение некоторого коэффициента $\left|\hat{F}(w)\right|$ меньше этой идеальной границы, теорема Парсеваля утверждает, что другие значения коэффициентов $\left|\hat{F}(w)\right|$ должны быть выше этой границы. Таким образом, попытка ограничить отдаленность абсолютных значений коэффициентов Уолша – Адамара от данной границы является возможным средством достижения высокой

нелинейности. Спектры некоторых функций содержат все значения (по модулю), равные этой идеальной границе. Такие функции называются бент-функциями.

Помимо обладания наивысшей возможной нелинейностью эти функции имеют нулевую автокорреляцию. Следовательно, функция стоимости

$$cost(f) = \sum_{\omega \in GF^n(2)} \left| \left| \hat{F}(\omega) \right| - 2^{\frac{n}{2}} \right|^R \tag{10}$$

является возможным подходом к оптимизации нелинейности и автокорреляции. Ввиду несбалансированности бент-функций приведенная функция стоимости *cost* может быть улучшена для нахождения сбалансированных криптографических функций. В [10] было введено обобщение функции стоимости (10), которое приняло следующий вид:

$$cost(f) = \sum_{\omega \in GF^n(2)} \left| \left| \hat{F}(\omega) \right| - X \right|^R . \tag{11}$$

Параметры X и R, называемые весовыми коэффициентами, обеспечивают свободу для экспериментирования и поиска оптимальных значений.

По аналогии с функциями стоимости относительно спектра Уолша – Адамара вида (11), функции стоимости относительно спектра автокорреляционной функции имеют следующий вид:

$$cost(f) = \sum_{s \in GF^n(2)} \left| r(s) \right| - X \Big|^R . \tag{12}$$

Традиционно ценовые функции применяются для оптимизации отдельной булевой функции. Для всего же нелинейного узла замен *cost* функции, основанные на спектре Уолша – Адамара, можно обобщить следующим образом [10]:

$$cost(f) = \sum_{\beta \in GF^m(2)} \sum_{\omega \in GF^n(2)} \left| \left| \hat{F}_\beta(\omega) \right| - X \right|^R \tag{13}$$

и аналогично для *cost* функций, основанных на автокорреляционном спектре:

$$cost(f) = \sum_{\beta \in GF^m(2)} \sum_{s \in GF^n(2)} \left| \left| r_\beta(s) \right| - X \right|^R . \tag{14}$$

Для оптимизации по критериям нелинейности и автокорреляции в [13] использовалась следующая функция стоимости:

$$cost(f) = \sum_{\beta \in GF^m(2)} \sum_{\omega \in GF^n(2)} \left| \left| \hat{F}_\beta(\omega) \right| - X_1 \right|^{R_1} + \sum_{\beta \in GF^m(2)} \sum_{s \in GF^n(2)} \left| \left| r_\beta(s) \right| - X_2 \right|^{R_2} . \tag{15}$$

В первой части проводимых в данной работе исследований использовались функции стоимости вида (12), (13), с заменой спектральных коэффициентов Уолша – Адамара и коэффициентов автокорреляционных спектров булевых функций на предложенные выше коэффициенты соответствующих спектров недвоичных функций.

Вторая часть проводимых исследований состояла в совершенствовании функций стоимости (критерия поиска криптографических функций), которое основывается на следующем положении. Известно, что при оптимизации криптографической функции по нелинейности и автокорреляции она по своим спектральным характеристикам (спектру корреляции с линейными функциями и автокорреляционному спектру) стремится к спектральным характеристи-

кам бент-функций, что и было использовано в предыдущих работах [10, 13] при разработке функций вида (10) – (15). В то же время очевидным недостатком такого подхода является использование одного (фиксированного) значения статического коэффициента, к которому стремятся все спектральные значения оптимизируемой криптографической функции. При этом значения спектральных коэффициентов идеальной функции (или бент-функции) состоят из двух возможных значений для булевых функций, и из трех значений для введенных недвоичных функций. При введении же дополнительных ограничений на сбалансированность количество возможных значений спектральных коэффициентов еще более возрастает.

При разработке новых функций стоимости в (11) – (15) предлагается заменить статический весовой коэффициент X на так называемые динамические весовые коэффициенты, т.е. весовые коэффициенты, принимающие различные значения для различных входных индексов спектра. В данной работе в качестве значений динамических весовых коэффициентов используются спектральные значения бент-функций. Предлагаемые функции стоимости имеют вид:

$$cost(f) = \sum_{\omega \in GF^n(2)} \left| \hat{F}(\omega) - \hat{B}(\omega) \right|^R, \tag{16}$$

$$cost(f) = \sum_{s \in GF^n(2)} \left| r_F(s) - r_B(s) \right|^R, \tag{17}$$

$$cost(f) = \sum_{\omega \in GF^n(2)} \left| \hat{F}(\omega) - \hat{B}(\omega) \right|^{R_1} + \sum_{s \in GF^n(2)} \left| r_F(s) - r_B(s) \right|^{R_2}, \tag{18}$$

где $\hat{B}(\omega), r_B(s)$ – спектральные значения нелинейности и автокорреляции случайной недвоичной бент-функции B.

В основе предлагаемого вычислительного метода синтеза регулярных нелинейных узлов замен симметричных криптоалгоритмов лежит применение математического аппарата недвоичных криптографических функций, методов корреляционного и спектрального анализа, а также предложенных в данной работе усовершенствованных ценовых функций (16) – (18) с использованием динамических весовых коэффициентов $\hat{B}(\omega), r_B(s)$. Усовершенствованный таким образом метод имитации отжига позволяет, как показано ниже, реализовать вычислительный поиск регулярных узлов замен с требуемыми показателями нелинейности и автокорреляции.

Результаты экспериментальных исследований

Для подтверждения достоверности и обоснованности полученных теоретических результатов проведены экспериментальные исследования эффективности предлагаемого вычислительного метода синтеза регулярных нелинейных узлов замен. Первая часть исследований проведена с использованием спектров недвоичных функций со статическими весовыми коэффициентами в функциях стоимости (13) – (15) метода имитации отжига, вторая часть исследований – с использованием динамических коэффициентов $\hat{B}(\omega), r_B(s)$ в функциях стоимости (16) – (18).

Параметры алгоритма для всех исследований были заданы следующими:

- $\alpha = 0.95$ – параметр геометрического охлаждения;
- $MIL = 500$ – число шагов, предпринимаемых во внутреннем цикле;
- $MaxIL = 300$ – максимальное число внутренних циклов поиска;
- $MUL = 50$ – максимальное число последовательных непродуктивных внутренних циклов;
- количество пробегов алгоритма для каждого набора параметров равно 10.

В ходе экспериментов с функциями стоимости вида (13) – (15) использовались различные значения статических коэффициентов X и фиксированное значение $R = 3$.

Формировались S-блоки размерностей 8x2, 4x4, 6x4. Узлы замен выходной размерности 4 представлялись через одну недвоичную функцию над $GF(2^4)$.

Полученные экспериментальные результаты для S-блоков 8x2 приведены в табл. 1. Лучший полученный результат выделен жирным шрифтом.

Таблица 1

Способ построения спектров, критерий отбора	Статические коэффициенты		Динамические коэффициенты	
	NL	*AC*	*NL*	*AC*
WHT, ср.	110	56	114	24
WHT, худш.	108	56	**116**	**24**
ACT, ср.	110	48	114	24
ACT, худш.	112	40	114	24
WHT+*ACT*, худш.	112	40	114	24
WHT+*ACT*, ср.	112	40	114	24

Как видно из полученных результатов, использование функций стоимости с динамическими весовыми коэффициентами позволило повысить показатели стойкости нелинейности и автокорреляции формируемых узлов замен.

В табл. 2 приведено сравнение полученных результатов с лучшими известными результатами, использующими традиционный подход описания S-блока в виде совокупности компонентных булевых функций [8 – 12]. Как видно из таблицы, полученные экспериментальные результаты предлагаемым методом с использованием функций стоимости на основе спектров недвоичных функций и статических весовых коэффициентов хорошо согласуются с экспериментальными результатами вычислительных методов традиционного подхода. Использование динамических весовых коэффициентов позволило получить лучшие известные на сегодняшний день результаты для S-блоков 8x2.

Таблица 2

Метод синтеза	*NL*	*AC*
Случайная генерация	108	56
Генетические алгоритмы	110	48
Имитация отжига (булевы функции)	114	32
Имитация отжига (недвоичные функции)	112	40
Имитация отжига (недвоичные функции, динамические весовые коэффициенты)	**116**	**24**

В табл. 3 приведены лучшие полученные результаты для S-блоков 4x4 и 6x4.

Таблица 3

S-блок	Метод имитации отжига	*NL*	*AC*
4x4	Булевы функции	4	8
4x4	Функции над $GF(2^4)$	4	8
6x4	Булевы функции	22	24
6x4	Функции над $GF(2^4)$	**24**	**24**

Как видно из приведенной таблицы, применение предлагаемого подхода позволяет повысить нелинейность формируемых S-блоков 6x4. Подобные S-блоки (размерности 6x4) применяются в DES-подобных шифрах.

Как показал анализ, S-блоки DES по своим криптографическим показателям нелинейности *NL* и автокорреляции *AC* далеки от оптимальных (см. данные для S1-S8 в табл. 4).

Таблица 4

S-блок	*NL*	*AC*	DDT_{max}
S1	14	48	16
S2	16	56	16
S3	16	48	16
S4	16	64	16
S5	12	**40**	**16**
S6	**18**	48	16
S7	14	52	16
S8	16	48	16
S1*-S8*	**24**	**24**	**10**

Разработанный вычислительный метод предлагается использовать для синтеза DES-подобных S-блоков с улучшенными криптографическими показателями стойкости (в табл. 4 приведены характеристики формируемых узлов замен S1*-S8*).

Следует отметить, что для обеспечения стойкости DES-подобных шифров к дифференциальному и линейному криптоанализу сформированные S-блоки необходимо оценивать не только по показателям нелинейности и автокорреляции, но и с учетом других критериев, учитывающих саму структуру шифра [14]. В этом смысле оценка эффективности формируемых S-блоков с учетом ограничений, накладываемых особенностями основных преобразований БСШ, а также апробация полученных результатов являются перспективным направлением дальнейших исследований.

Заключение

Предложенная математическая модель представления S-блоков через недвоичные функции является новым направлением в области формирования нелинейных узлов замен и представляет собой основу для дальнейших исследований в этом направлении. Введенное теоретическое обобщение и предложенный математический аппарат недвоичных криптографических функций позволил усовершенствовать вычислительный метод формирования нелинейных узлов замен (метод имитации отжига). В результате процесс формирования нелинейных узлов замен с требуемыми криптографическими свойствами упрощен, а именно – существенно сократилось количество оптимизационных спектров в процессе синтеза S-блоков.

Усовершенствованный вычислительный метод реализован программно, полученные экспериментальные результаты с использованием функций стоимости со статическими весовыми коэффициентами хорошо согласуются с экспериментальными результатами вычислительных методов традиционного подхода. Использование динамических весовых коэффициентов в новых разработанных функциях стоимости позволили получить лучшие известные на сегодняшний день результаты для S-блоков 8x2. Для S-блоков 6x4 удалось поднять верхнюю границу показателя нелинейности. Разработанный вычислительный метод предлагается использовать для синтеза DES-подобных S-блоков.

Перспективными направлениями дальнейших исследований являются развитие математического аппарата криптографических недвоичных функций для задач синтеза S-блоков, экспериментальные исследования эффективности предлагаемого подхода для узлов замен больших размерностей, обобщение динамических весовых коэффициентов.

Список литературы: 1. *Сорока, Л.С.* Вероятностная модель формирования нелинейных узлов замен для симметричных криптографических средств защиты информации / Сорока Л.С., Кузнецов А.А., Московченко И.В., Исаев С.А. // Системи обробки інформації. – Х. :ХУВС, 2009. – № 3 (77). – С. 101-104. 2. *O'Connor, L.* An analysis of a class of algorithms for S-box construction / O'Connor L. // J. Cryptology. - 1994. – P. 133-151. 3. *Сорока, Л.С.* Исследование вероятностных методов формирования нели-

нейных узлов замен / Сорока Л.С., Кузнецов А.А., Исаев С.А. // Системи обробки інформації. – 2011. – № 8 (98). – С. 113 – 122. 4. *Булева функция* // Режим доступу: http://ru.wikipedia.org/wiki/Булева_функция. 5. *Dawson, E.* Designing Boolean functions for cryptographic applications / Dawson E., Millan W., Simpson L. // Contributions to General Algebra, Verlag Johannes Heyn, Klagenfurt. – 2000. – 12. – P. 1-22. 6. *Clark, J.A.* Evolving Boolean functions satisfying multiple criteria / Clark J.A., Jacob J.L., Stepney S., Maitra S., Milan W. // Lecture Notes in Computer Science (2551), Springer, Berlin. – 2002. – 2251. – P. 246-259. 7. *Parker, M.G.* Generalised S-Box Nonlinearity / Parker M.G. // NES/DOC/UIB/WP5/020/A. – 2003. 8. *Millan, W.* How to improve the nonlinearity of bijective s-boxes / Millan W. // Information Security and Privacy, ACISP '98, Springer Verlag. – 1998. – volume 1438 of Lecture Notes in Computer Science. – P. 181-192. 9. *Millan, W.* Evolutionary Heuristics for Finding Cryptographically Strong S-Boxes / Millan W., Burnett L., Carter G., Clark A., Dawson E. // Information and communication security, Springer, Heidelberg. – 1999. – Lecture Notes in Computer Science Volume 1726. – P.263-274. 10. *Clark, J.A.* The Design of S-Boxes by Simulated Annealing / Clark J.A., Jacob J.L., Stepney S. // New Generation Computing. – 2005. – 23(3). – P.219–231. 11. *Laskari, C.* Utilizing Evolutionary Computation Methods for the Design of S-Boxes / Laskari C., Meletiou C., Vrahatis N. // Computational Intelligence and Security. – 2006. – Volume 2. – P.1299-1302. 12. *Tesar, P.* A new method for generating high non-linearity S-Boxes / Tesar P. // Radioengineering. – 2010. – Part I of II, Vol. 19 Issue 1. – P.23 -26. 13. *Kavut, S.* Improved Cost Function in the Design of Boolean Functions Satisfying Multiple Criteria / Kavut S., Yücel M.D. // Proc. INDOCRYPT. – 2003. – P.121-134. 14. *Kwangjo, K.* Securing DES S-Boxes against Three Robust Cryptanalysis / Kwangjo K., Sangjin L., Sangjoon P., Daiki L. // Proceedings of the Workshop on Selected Areas in Cryptography, SAC '95. – 1995. – P.145-157.

Харьковский национальный
университет радиоэлектроники

Поступила в редколлегию 14.09.2012

КРИПТОГРАФИЧЕСКИЕ ПРЕОБРАЗОВАНИЯ

УДК 004 056 55

Е.Г. КАЧКО, канд. техн. наук, Д.С. БАЛАГУРА, канд. техн. наук,
К.А. ПОГРЕБНЯК, канд. техн. наук, Ю.И. ГОРБЕНКО, канд. техн. наук

ИССЛЕДОВАНИЕ МЕТОДОВ ВЫЧИСЛЕНИЯ ИНВЕРСИИ В АЛГОРИТМЕ NTRU

Операция инверсии используется для генерации личных ключей шифрования и цифровой подписи и является вычислительно сложной операцией. Цель данной работы - исследование различных методов вычисления инверсии, их оптимизация с учетом возможностей параллельных вычислений. Исследуются алгоритмы [1 – 3] с точки зрения их возможной параллелизации и сравниваются по вычислительной сложности.

В алгоритме NTRU используется кольцо усеченных полиномов [1] $Z(X)/(X^N-1)$, которые содержат не более чем N целочисленных элементов. Значения коэффициентов полинома обычно ограничены. Кольцо полиномов, коэффициенты которых меньше p, обозначается как: $(Z/pZ)[X]/(X^N-1)$. Принято обозначать полиномы, принадлежащие такому кольцу, как $a(X), b(X),...$ Инверсия полинома $a(X)$ в кольце $(Z/pZ)[X]/(X^N-1)$ – операция определения полинома $b(X)$, такого, что $a(X)*b(X)=1$ в этом же кольце.

Рассмотрим алгоритм инверсии, предложенный в [1]

Алгоритм 1 (А1). Вычисление инверсии.

Вход. Полином $a(X)$, порядок полинома N, модуль q [1]

Выход. Статус (Успех или ошибка), обратный элемент $b(X)$ (в случае успешного статуса).

1. Приведение $a(X)$ в элемент кольца $(Z/2Z)[X]/(X^N-1)$ (все коэффициенты полинома приводятся по модулю $p=2$, отрицательные значения заменяются положительными $(2+a[i]\%2)$.
2. Используя расширенную теорему Эвклида для полиномов $a(X)$ и X^N-1, находим наибольший общий делитель $d(X)$, и коэффициенты $u(X)$, $v(X)$ Диофантового уравнения $u(X)*a(X)+v(X)*(X^N-1)=d(X)$ для модуля $p=2$.
3. Если наибольший общий делитель $d(X)$ не равен 1 ($d[0]\neq\pm1$ или $d[1],d[2],...,d[N-1]\neq 0$) – обратного элемента нет, статус равен *Ошибка* и *Выход*.
4. Вычислить $b(X)=(d(X)^{-1}(\bmod p)*u(X))\bmod p$.
5. Для всех модулей p = 4, 16, 256, 65536 ($p=p^2$).

5.1. Приводим $a(X)$ по модулю p (отрицательные значения заменяем значениями $p + a[i]$).

5.2 Вычисляем новое значение $b(X)=(2*b(X)-a(X)*b(X)^2)\bmod p$.

6. Значение *b(x)* приводим по модулю q = 2048 .
7. Статус равен *Успешно.*
8. Выход.

Алгоритмы, рассмотренные далее (А2, А3, А4), предназначены для вычисления инверсии для модуля 2 (пункты 2 – 4 алгоритма А1).

---

[1] Для всех полиномов в [1] q = 2048.

Алгоритм 2 (А2)

Вход. Порядок p[2], полиномы $a(X)$, $c(X)$[3].

Выход. Полиномы $d(X), u(X), v(X)$, удовлетворяющие уравнению: $a(X)*u(X)+c(X)*v(X)=d(X)$ в кольце.

1. Если $c(X)$, то $u(X)=1$, $v(X)=0$, $d(X)=a(X)$.
2. $u(X); d(X)=a(X); v_1(X)=0; v_3(X)=c(X)$.
3. Пока ($v_3(X)\neq 0$) выполнить:

3.1. $q(X)=d(X)\%v_3(X)$[4]; $t_3(X)=d(X)\%v_3(X)$[5];

3.2. $t_1(X)=u(X)-q(X)*v_1(X)$;

3.3. $u(X)=v_1(X)$;

3.4. $d(X)=v_3(X)$;

3.5. $v_1(X)=t_1(X)$;

3.6. $v_3(X)=t_3(X)$;

4. $v(X)=(d(X)-a(X)*u(X))/b(X)$.

Рассмотрим оптимизацию А2 как составной части А1.

1. Полином $c(X)=X^N-1$ не может быть равным 0, поэтому первый пункт алгоритма можно опустить.
2. При выполнении первой итерации цикла в операции деления $a(X)/c(X)$ отметим что порядок $a(X)$ меньше порядка $c(X)$, следовательно, на первом шаге $q(X)=0$, $t_3(X)=a(X)$. Таким образом, первую итерацию можно опустить целиком. Для второй итерации использовать деление $c(X)/a(X)$, а в качестве начального значения $d(X)$ использовать значение $a(X)$.
3. При вычислении $t_1(X)$ после очередного деления $u(X)=0$, а $v_1(X)=1$, т.е. $t_1(X)=-q(X)$, а с учетом модуля 2 $t_1(X)=q(X)$.
4. Так как все коэффициенты и результат вычисления по модулю 2, операцию умножения (шаг 3.2) следует заменить операциями сложения по модулю 2.
5. Значение $v(X)$ не используется далее, его вычисление можем опустить (пункт 4 алгоритма 2).
6. Везде, где возможно, следует использовать SSE операции.

Алгоритм 2 после оптимизации (А2'):

1. $d(X)=a(X); u(X)=1$; $q(X)=c(X)/a(X); v_3(X)=c(X)\%a(X)$.
2. $v_1(X)=q(X)$.
3. Пока ($v_3(X)\neq 0$) выполнить:

3.1. $q(X)=d(X)/v_3(X)$; $t_3(X)=d(X)\%_3(X)$;

3.2. $t_1(X)=u(X)\hat{}q(X)*v_1(X)$; $u(X)=v_1(X)$, $d(X)=v_3(X)$;

3.4. $v_1(X)=t_1(X)$; $v_3(X)=t_3(X)$.

В одной строчке задаются операторы, которые можно выполнять параллельно.

Результаты, полученные после предложенной оптимизации, приведены в табл.1.

---

[2] Значение p=2.

[3] Последнему полиному соответствует полином $x^N - 1$.

[4] Деление нацело, дробная часть отбрасывается.

[5] Вычисление остатка от деления.

Как видно из таблицы, оптимизация алгоритма А2 позволила ускорить его не менее, чем в четыре раза.

Таблица 1

Оптимизация расширенного алгоритма Эвклида для NTRU[6]

Порядок полинома N	Время, *мс*		Ускорение, S=t1/t2
	A2, t1	A2', t2	
401	4,04	0,826	4,89
449	5,17	0,997	5,19
677	11,56	2,18	5,30
1087	22,99	4,75	4,84
541	5,82	1,38	4,22
613	6,75	1,51	4,47
887	14,49	3,03	4,78
1171	25,76	5,41	4,76
659	8,08	1,99	4,06
761	10,71	2,46	4,35
1087	21,21	4,71	4,50
1499	39,25	8,85	4,44

Продолжим оптимизацию Алгоритма 1. Шаг 3 проверяет, что наибольший общий делитель равен 1 (-1). Заметим, что полином $x^N - 1$. Он имеет корень $X = 1$, который не может быть корнем полинома $f(X)$, используемого для вычисления личного ключа. Действительно, $f(X) = 3 * F(X) + 1$, полином $F(X)$ имеет одинаковое число коэффициентов равных 1 и -1, следовательно, $X = 1$ – корень полинома $F(X)$, т.е. не является корнем $f(X)$. Второй множитель $x^N - 1$ имеет порядок N, как и полином $f(X)$, но не совпадает с ним, так как все его коэффициенты равны 1. Для полинома $g(X)$ количество 1 и -1 отличается на 1, что обеспечивает отсутствие корня $X = 1$. Следовательно, проверка наибольшего общего делителя не имеет смысла и может быть опущена. Шаг 4, в котором выполняется вычисление по формуле $b(X) = (d(X)^{-1} (\mathrm{mod}\, p) * u(X)) \bmod p$, фактически также может быть опущен, так как $d(X) = 1$, обратный элемент для единичного элемента равен 1, а вычисления $u(X)$ изначально выполняются по модулю 2.

Таким образом, алгоритм 2 в улучшенном варианте можно рассматривать не только как реализацию расширенного алгоритма Эвклида, но и как вычисление обратного элемента по модулю 2.

Использование almost inverses (почти инверсий)[2]

При вычислении инверсии по модулю 2 предыдущим алгоритмом использовались операция деления, в результате которой порядок полинома на каждом шаге уменьшался. Вместо обнуления старших коэффициентов полинома при вычислении инверсии предложено обнуление младших коэффициентов [2] и вычисление общего числа таких коэффициентов k. В этом случае фактически решается уравнение $a(X) * u(X) + c(X) * v(X) = X^k$. Значение $u(X)$ авторы [2] называют Almost Inverses (почти инверсия) и предлагают использовать этот алгоритм для вычисления инверсии для эллиптических кривых. Для получения значения инверсии результат достаточно разделить на X^k.

[6] Результаты приведены для порядков полинома, рекомендованных [1].

В работе [3] показана применимость этого метода для вычисления инверсии по модулю 2 в алгоритме NTRU. В качестве необходимых условий возможности применения этого алгоритма используются условия:

$$\gcd(a(X), c(X)) = 1;\ c(0) = 1.$$

Как показано выше, данные условия выполняются.

Алгоритм 3 (А3) [3]

Вход. Порядок $p = 2$, полиномы $a(X)$; $c(X) = X^N - 1$.

Выход. Полином $b(X)$, такой что $b(X) = a(X)^{-1}$ в кольце $(Z/2Z)/[X]/(X^N - 1)$.

1. $k = 0$; $b(X) = 1$; $c(X) = 0$; $f(X) = a(X)$; $g(X) = x^N - 1$.
2. Выполнить цикл:

2.1. Определение количества младших нулевых коэффициентов (r) в $f(X)$;

2.2. Сдвиг $f(X)$ вправо на r элементов;

2.3.Сдвиг $c(X)$ влево на r элементов;

2.4. $k+ = r$;

2.5.Если $f(X) == 1$ то Выход из цикла;

2.6.Если порядок $f(X)$ меньше порядка $g(X)$, то поменять местами $f(X)$ и $g(X)$, $b(X)$ и $c(X)$;

2.7. $f(X) = f(X)\hat{}\, g(X)$; $b(X) = b(X)\hat{}\, c(X)$.

3. $b(X) = b(X) * X^{(N-k)\%N}$.

Оптимизация алгоритма А3.

1. Вместо замены местами полиномов использовать замену местами адресов этих полиномов.
2. Для выполнения этапа 3 использовать циклический сдвиг полинома.
3. Везде, где возможно, использовать SSE команды

Результаты сравнения алгоритмов А2' и оптимизированного А3 будут приведены ниже после рассмотрения очередного алгоритма.

Использование улучшенного алгоритма вычисления GCD и обратного элемента [5]

Этот алгоритм отличается от алгоритма А2 тем, что вычисление выполняется с помощью прямого и обратного преобразований. При прямом преобразовании выполняется последовательное вычисление частного и остатка, как в алгоритме вычисления GCD. При этом все частные запоминаются в стеке. При обратном преобразовании выполняется фактическое вычисление обратного элемента, если GCD = 1. В противном случае делается вывод об отсутствии обратного элемента.

Алгоритм вычисления обратного элемента с учетом идеи автора [5], в применении к полиномам:

Алгоритм А4.

Прямой ход:

1. $d(X) = c(X); v_3(x) = a(x); r = 0$.
2. Пока ($v_3(X) \neq 1$) выполнить:

2.1. $q(X) = d(X)/v_3(X)$; $t_3(X) = d(X)\%v_3(X)$;

2.2. push ($q(x)$) r++;

2.3. $d(x) = v_3(x)$;

2.4. $v_3(x) = t_3(x)$.

Обратный ход

1. $S(x) = 0$;
2. Для $i = 1, 2, \ldots, r$ выполнить:

2.1. $q(x) = pop()$;

2.2. $b(x) = S(X) + q(x) * v_3(x)$;

2.3. $S(x) = v_3(x)$;

2.4. $v_3(x) = b(x)$.

Таблица 2

Сравнение всех алгоритмов вычисления инверсии по модулю 2

Порядок полинома N	Время, *мс*			Ускорение, S1=t2/t3	Ускорение, S1=t2/t4	Ускорение, S3=t4/t3
	A2'(t2)	A3(t3)	A4 (t4)			
401	0,826	0,25	0,80	3,30	1,03	3,20
449	0,997	0,32	0,96	3,12	1,04	3,00
677	2,18	0,63	1,9987	3,46	1,09	3,17
1087	4,75	1,62	4,63	2,93	1,03	2,86
541	1,38	0,41	1,38	3,37	1,00	3,37
613	1,51	0,53	1,47	2,85	1,03	2,77
887	3,03	1,04	3,00	2,91	1,01	2,88
1171	5,41	1,96	5,44	2,76	0,99	2,78
659	1,99	0,61	1,83	3,26	1,09	3,00
761	2,46	0,82	2,43	3,00	1,01	2,96
1087	4,71	1,61	5,12	2,93	0,92	3,18
1499	8,85	3,19	-	2,77	-	-

Теоретический анализ производительности алгоритма, проведенный автором [5], обещал улучшение производительности приблизительно на 54 %, но при этом не учитывались свойства кэша. В случае полинома элемент кэша – это полином, число элементов кэша приблизительно равно N/2. Таким образом, весь стек во внутреннем кэше не помещается, это приводит к существенным потерям производительности. Более того, для N = 1499 необходимо вместо выделения памяти в стеке использовать выделение динамической памяти, что также замедляет выполнение функции. В табл. 2 приведены результаты сравнения трех оптимизированных алгоритмов вычисления обратного элемента для модуля, равного 2^7.

[7] Значение N=1499 для алгоритма A4 не используется.

Таким образом, использование алгоритма А3 вместо оптимизированного алгоритма А2' позволяет увеличить производительность еще примерно в три раза, получить общее увеличение производительности не менее, чем в 12 раз. Использование же усовершенствованного алгоритма вычисления инверсии (А4) по сравнению с обычным (А2), если и дает выигрыш, то максимум на 9 %, а алгоритм вычисления Almost Invers (А3) существенно выигрывает.

Заметим, что полученные и результаты противоречат результатам из [4 и 5]. Это связано с тем, что в указанных работах определялось только общее число необходимых итераций и не учитывались возможности их параллельного выполнения, а также свойства кэша.

Вернемся к оптимизации алгоритма А1.

При вычислении по формуле $b(X) = (2 * b(X) - a(X) * b(X)^2) \bmod p$ для вычисления $b(x)^2$ используется формула сокращенного умножения, а операции сложения, вычитания и приведения по модулю выполняются с помощью SSE команд.

В табл. 3 приведены окончательные результаты, полученные после оптимизации алгоритма А1 (алгоритм А1'), а на рисунке – диаграмма с полученными результатами.

Таблица 3

Вычисление инверсии по модулю $p = 2048$

Порядок полинома N	Время, *мс*		Ускорение, S=t1/t2
	A1, t1	Улучшенный A1, t2	
401	19,42	3,89	4,99
449	23,89	5,07	4,71
677	48,60	9,73	4,99
1087	83,56	20,10	4,16
541	19,06	4,37	4,36
613	23,70	5,52	4,29
887	51,95	11,56	4,49
1171	90,78	20,29	4,47
659	25,75	5,75	4,48
761	33,26	7,53	4,42
1087	67,96	15,76	4,31
1499	124,85	28,99	4,31

В алгоритме А' для вычисления инверсии по модулю 2 используется алгоритм А3. Для параллелизации вычислений используются SIMD команды. Потоки не используются.

Заметим, что ускорение распределено практически равномерно; это связано с тем, что системные погрешности оказывают на этот алгоритм меньше влияние, чем на предыдущие алгоритмы, так как время выполнения этого алгоритма больше.

Таким образом, оптимизация алгоритма вычисления инверсии привела к более чем четырехкратному увеличению скорости

Список литературы: 1. *American* National Standard for Financial Services ANSI X9.98 – 2010. Lattice-Based Polynomial Public Key Establishment Algorithm for the Financial Services Industry. 2. *R. Schroeppel, H. Orman, S. O'Malley and O. Spatscheck*, "Fast key exchange with elliptic curve systems".Advances in Cryptology, Proc. Crypto'95, LNCS 963. 3 NTRU Cryptosystems Technical Report.. http://www.securityinnovation.com/uploads/Crypto/NTRUTech014.pdf. 4. *Kyle Wilhelm*. Aspects of Hardware Methodologies for the NTRU Public-Key Cryptosystem. https://ritdml.rit.edu/bitstream/handle/1850/7774/KWillaimsThesis02-2008.pdf?sequence=1. 5. *Boris S. Verkhovsky*. Enhanced Euclid Algorithm for Modular Multiplicative Inverse and Its Application in Cryptographic Protocols. *Int. J. Communications, Network and System Sciences*, 2010, 3, 901-906.

Харьковский национальный университет радиоэлектроники

Поступила в редколлегию 20.09.2012

УДК 681.3.06

Ю.М. ЛЄНШИНА, канд. техн. наук, А.В. ЛЄНШИН, канд. техн. наук

КРИПТОГРАФІЧНА ПІДТРИМКА ПОСЛУГ ЗАХИСТУ ВІД НЕСАНКЦІОНОВАНОГО ДОСТУПУ

Вступ

Несанкціонованим доступом (НСД) до інформаційних ресурсів називається, доступ здійснений з порушенням встановлених політикою безпеки правил доступу користувачів та процесів до пасивних об'єктів. Отже завданням діяльності із захисту від НСД є забезпечення дотримання правил доступу, що встановлені політикою безпекою. У цьому випадку, під політикою безпекою розуміють сукупність правил, які поділяють стани системи, у якій впроваджуються механізми захисту, на підмножини «дозволені стани» та «недозволені стани».

Захист інформації від НСД полягає в створенні і підтримці в дієздатному стані системи заходів та засобів захисту інформації, як технічних (інженерних, програмно-апаратних), так і нетехнічних (правових, організаційних), що запобігають або ускладнюють реалізацію загроз, знижують очікувані збитки. При цьому розглядаються як суб'єктивні (джерелом є людина), так і об'єктивні загрози, що є наслідками дій сил природи чи відмов обладнання [1].

Основними видами доступу, що можуть бути здійснені суб'єктами доступу (користувачами або процесами) відносно пасивних об'єктів є такі як: читання, модифікація, запуск на виконання. Керування цими видами доступу дозволяє забезпечувати такі властивості безпеки інформації, як конфіденційність, цілісність, доступність та спостереженість. Стрімкий розвиток інформаційних технологій призвів до виникнення ще однієї проблеми захисту інформації – атак, спрямованих на несанкціоноване отримання ідентифікаційних даних користувачів інформаційних систем. Для протидії зазначеним атакам на практиці реалізуються послуги приватності: анонімність, псевдонімність, прихованість та неможливість асоціації.

Незважаючи на різноманітність підходів до побудови систем захисту найчастіше використовується підхід, що передбачає забезпечення достатнього рівня захисту за мінімальних затрат на впровадження та супровід комплексної системи захисту інформації. Таким чином, актуальним є питання кількісної, або принаймні, якісної оцінки стійкості послуг захисту від НСД, що забезпечується комплексом засобів захисту (КЗЗ).

У статті досліджуються підходи, які за рахунок застосування методів криптографічної підтримки, дозволяють протидіяти загрозам несанкціонованого доступу, зокрема, загрозам порушення приватності.

Аналіз вимог національних нормативних документів в сфері захисту від НСД

У 1999 році в Україні введено в дію перші документи з серії нормативних документів з технічного захисту інформації, що визначають вимоги та порядок захисту інформації від НСД, відомих під назвою НД ТЗІ [1]. Незважаючи на те, що у наступні роки 2000 – 2012 уповноваженими органами у сфері захисту інформації України (Департаментом спеціальних телекомунікаційних систем та захисту інформації Служби безпеки України та його правонаступниками) розроблені інші документи, їх скоріше слід вважати тлумаченням окремих вимог, що були встановлені у 1999 році.

Основним з НД ТЗІ, що визначає критерії оцінки захищеності комп'ютерної системи (КС) від НСД є НД ТЗІ 2.5-004-99 [1]. У вітчизняній літературі під КС розуміється сукупність програмно-апаратних засобів, яка подана для оцінки. Тобто фактично КС є аналогом об'єкта оцінки (ТОЕ – Target Of Evaluation) за ISO/IEC 15408 [2], відомого під назвою "Єдині критерії". Встановлені у НД ТЗІ 2.5-004-99 критерії дають змогу системно розв'язувати завдання проектування КЗЗ від НСД у окремих продуктах, створення комплексних систем захисту інформації, а також проведення оцінювання відповідності розроблених КС у результаті проведень незалежних оцінювань.

Сукупність вимог (шкала оцінки), що закріплена у НД ТЗІ 2.5-004-99 застосовується для оцінювання ефективності функціональних послуг безпеки і коректності їх реалізації. Вимоги закріплені у вигляді функціональних критеріїв та критеріїв гарантій.

Функціональні критерії згруповані за ознакою властивості безпеки, для якої призначено відповідні послуги. Функціональним критерієм чи послугою безпеки, називаються правила, згідно з якими функціонують механізми, що реалізують послугу.

Критерії гарантій включають вимоги до архітектури КЗЗ, середовища розробки, послідовності розробки, випробування, середовища функціонування і експлуатаційної документації. У НД ТЗІ 2.5-004-99 вводиться сім рівнів гарантій, що є ієрархічними. Ієрархія рівнів гарантій відбиває поступово зростаючу міру впевненості, що послуги, які надаються, дозволяють протистояти загрозам, а механізми, що їх реалізують – коректно реалізовані, і можуть забезпечити очікуваний споживачем рівень захищеності інформації під час експлуатації КС.

Як випливає із аналізу змісту послуг безпеки, задача забезпечення послуги приватності у системі НД ТЗІ не розглядається. Такий стан скоріше за все зумовлений тим, що зазначена серія НД ТЗІ грунтується на принципах та правилах визначених у документах, зокрема Канадських критеріях (1993) [3], що були розроблені ще до того як проблема забезпечення приватності набула актуальності. Таким чином, можна стверджувати, що вітчизняна нормативна база на сьогодні не надає критеріїв проектування та оцінювання рішень, що забезпечують захист від загроз порушення приватності.

Аналіз вимог міжнародного стандарту ISO/IEC 15408

Міжнародний стандарт ISO/IEC 15408 відомий також під назвою "Єдині критерії" визначає вимоги до проектування, оцінювання функцій безпеки та рівнів гарантій, що реалізуються КЗЗ. ISO/IEC 15408, як і НД ТЗІ, був розроблений на основі "Канадських критеріїв" та "Федеральних критеріїв США".

Структурно ISO/IEC 15408 [2] складається з трьох частин, що визначають відповідно: базові вимоги до проектування оцінювання функцій захисту, реалізованих у ІТ-продуктах, каталог функціональних вимог безпеки та каталог вимог довіри (гарантій). З метою систематизації вимог стандарту, усі вимоги безпеки та гарантій стандарту згруповані за ієрархічним принципом: клас-родина-компонент-елемент. Класи визначають найбільш загальне групування вимог. Родини у межах класу розрізняються за ступенем суворості та іншим характеристикам вимог. Компонент є тим мінімальним набором вимог, що може бути включений до завдання з безпеки (аналог технічного завдання на КЗЗ) або профілю захисту (аналог технічного завдання на "типовий" КЗЗ). У якості елементу виступає неподільна вимога.

Між компонентами визначені залежності. Залежності виникають у випадках, коли компонент не є самодостатнім, і отже при застосуваннях компоненту, що залежить від інших, усі залежності мають бути задоволені шлях включення до технічного завдання усіх залежних компонентів.

Слід зауважити, що хоча вимоги ISO/IEC 15408 та НД ТЗІ згруповані за ієрархічним принципом, але ознаки для класифікації в них використовуються різні, з цього випливає, що послуга безпеки як основний елемент НД ТЗІ, складається з кількох компонентів ISO/IEC 15408 [2]. Класами вимог безпеки, що визначені у ISO/IEC 15408 є: аудит безпеки (FAU), зв'язок (FCO), криптографічна підтримка (FCS), захист даних користувача (FDP), ідентифікація та автентифікація (FIA), керування безпекою (FMT), приватність (FPR), захист функцій безпеки об'єкту оцінювання (FPT), використання ресурсів (FRU), доступ до об'єкту оцінювання (FTA), довірений маршрут/канал (FTP).

Проведений порівняльний аналіз змісту вимог безпеки з ISO/IEC 15408 та НД ТЗІ дозволяє стверджувати, що вони близькі за своїм наповненням, хоча ISO/IEC 15408 надає більшу гнучкість при виборі функцій безпеки, а в НД ТЗІ, нажаль, відсутні вимоги, що викликають підвищений інтерес з огляду на тему дослідження, а саме це питання: крипто-

графічної підтримки (клас FCS) та надання послуг з приватності (клас FPR).

Починаючи з першої редакції ISO/IEC 15408 [2] у ньому визначаються вимоги із забезпечення приватності. Основною задачею забезпечення приватності є: захист від розкриття ідентифікаційних даних користувача та зловживання ним з боку інших користувачів.

З метою кращого розуміння сутності вимог із забезпечення приватності у ході дослідження побудовано інформаційно-понятійну модель (рис. 1).

За результатами аналізу змісту вимог із забезпечення приватності, структури НД ТЗІ та ISO/IEC 15408 у роботі були сформульовані пропозиції з розширення функціональних послуг безпеки, що визначаються у НД ТЗІ, за рахунок критеріїв приватності. Беручи до уваги систему позначень, прийняту у НД ТЗІ, було вирішено застосовувати такі скорочення для функціональних послуг безпеки: "FPR_ANO" – "ПА", "FPR_PSE" – "ПП", "FPR_UNL" – "ПН", "FPR_UNO" – "ПР", у якості розділювача між скороченням для послуги та її рівнем за аналогією до НД ТЗІ було використано знак дефіса "-".

У табл. 1 – 3 наведені рекомендовані форми специфікації послуг "Анонімність" та "Псевдонімність", "Неможливість асоціації".

Таблиця 1

ПА-1 Анонімність	ПА-2 Анонімність без запиту інформації
КЗЗ має забезпечити, щоб [призначення: користувачі та/або суб'єкти] були не здатні визначити справжнє ім'я користувача, що пов'язаний з [призначення: суб'єкти та/або операції, та/або об'єкти]	
—	КЗЗ має надавати [призначення: послуги] для [призначення: суб'єкти] без запиту будь-якого посилання на справжнє ім'я користувача
НЕОБХІДНІ УМОВИ: НЕМАЄ	

Таблиця 2

ПП-1 Псевдонімність	ПП-2 Реверсивна псевдонімність	ПП-3 Альтернативна псевдонімність
КЗЗ має забезпечити, щоб [призначення: користувачі або суб'єкти] були не здатні визначити справжнє ім'я користувача, що пов'язаний з [призначення: суб'єкти та/або операції, та/або об'єкти]		
КЗЗ має бути здатний надати [призначення: кількість псевдонімів] псевдонімів справжнього імені користувач для [призначення: перелік суб'єктів]		
КЗЗ має бути здатний [вибір (обрати одне з): визначити псевдонім користувача, прийняти псевдонім від користувача] та перевірити його на відповідність [призначення: метрика псевдонімів]		
—	КЗЗ має надати [вибір: вповноважений користувач, [призначення: довірені суб'єкти]] можливість визначати ідентифікаційні дані користувача за наданим псевдонімом тільки за умови [призначення: умови]	КЗЗ має надати псевдонім для справжнього імені користувача, який має бути ідентичний псевдоніму наданому раніше за наступних умов [призначення: умови]; інакше наданий псевдонім має бути не пов'язаний з наданими раніше псевдонімами
—	НЕОБХІДНІ УМОВИ: НИ-1	—

Таблиця 3

ПН-1 Неможливість асоціації
КЗЗ має забезпечити, щоб [призначення: користувачів та/або суб'єкти] були не здатні визначити, що [призначення: перелік операції] [вибір: були ініційовані одним і тим самим користувачем, пов'язані наступним чином [призначення: перелік співвідношень]]
НЕОБХІДНІ УМОВИ: НЕМАЄ

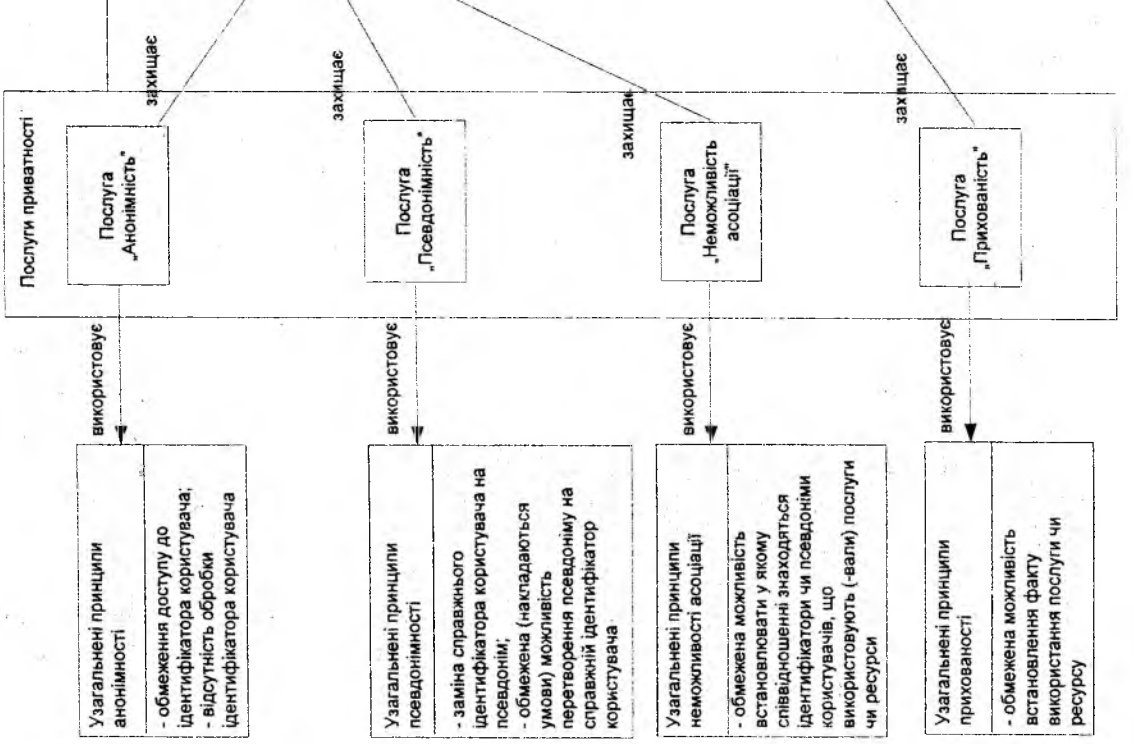
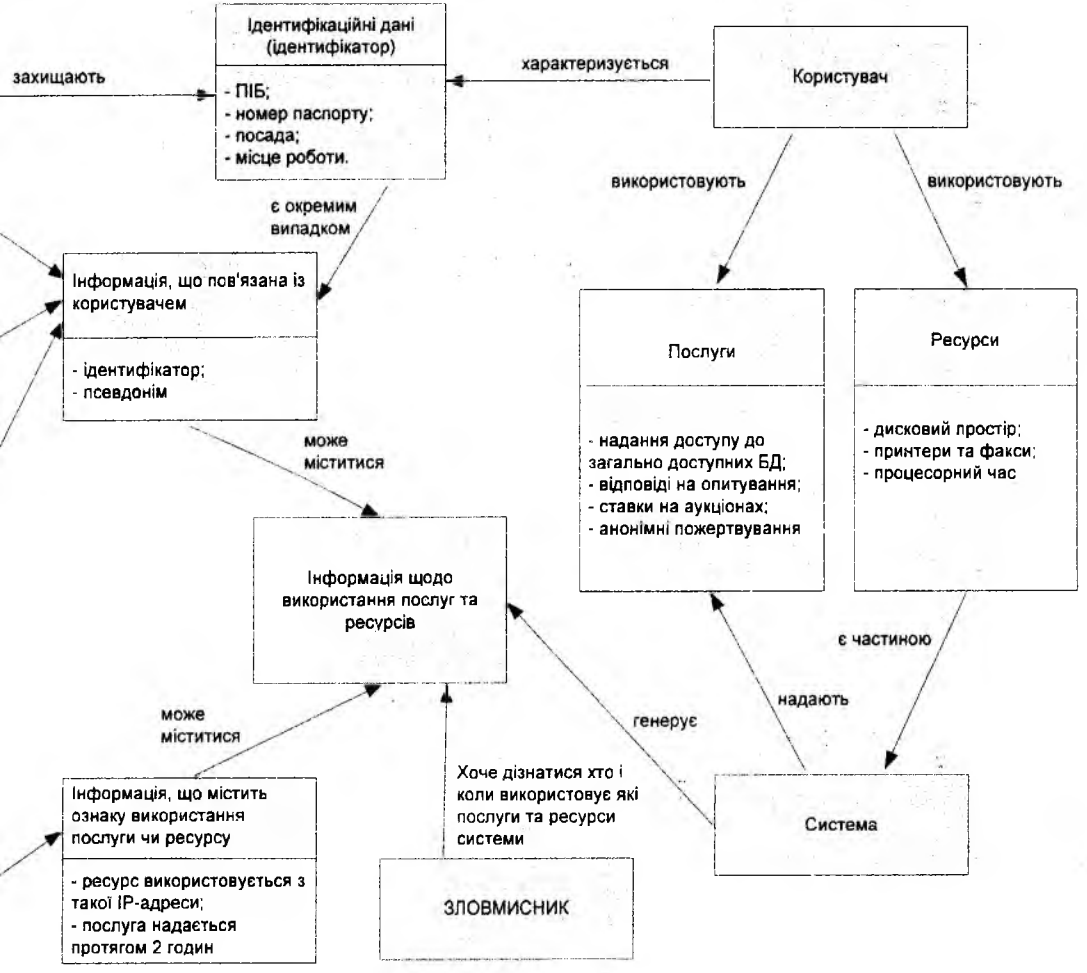

Послуги приватності
Послуга „Анонімність"
Послуга „Псевдонімність"
Послуга „Неможливість асоціації"
Послуга „Прихованість"
захищає
захищає
захищає
захищає
використовує
використовує
використовує
використовує
Узагальнені принципи анонімності
- обмеження доступу до ідентифікатора користувача;
- відсутність обробки ідентифікатора користувача
Узагальнені принципи псевдонімності
- заміна справжнього ідентифікатора користувача на псевдонім;
- обмежена (накладаються умови) можливість перетворення псевдоніму на справжній ідентифікатор користувача
Узагальнені принципи неможливості асоціації
- обмежена можливість встановлювати у якому співвідношенні знаходяться ідентифікатори чи псевдоніми користувачів, що використовують (-вали) послуги чи ресурси
Узагальнені принципи прихованості
- обмежена можливість встановлення факту використання послуги чи ресурсу

Рис.1


Ідентифікаційні дані (ідентифікатор)
- ПІБ;
- номер паспорту;
- посада;
- місце роботи.
захищають
характеризується
Користувач
використовують
використовують
є окремим випадком
Інформація, що пов'язана із користувачем
- ідентифікатор;
- псевдонім
Послуги
- надання доступу до загально доступних БД;
- відповіді на опитування;
- ставки на аукціонах;
- анонімні пожертвування
Ресурси
- дисковий простір;
- принтери та факси;
- процесорний час
може міститися
Інформація щодо використання послуг та ресурсів
є частиною
надають
може міститися
генерує
Інформація, що містить ознаку використання послуги чи ресурсу
- ресурс використовується з такої IP-адреси;
- послуга надається протягом 2 годин
Хоче дізнатися хто і коли використовує які послуги та ресурси системи
Система
ЗЛОВМИСНИК

Криптографічна підтримка та оцінювання стійкості послуг захисту від НСД

У ISO/IEC 15408 визначається, що КЗЗ може використовувати криптографічні функціональні можливості для сприяння досягненню деяких найбільш важливих цілей безпеки. Зауважується, що до них відносяться, але не обмежуються, такі цілі як: ідентифікація, неспростовність, довірений маршрут та канал, розподіл даних. Таким чином, ISO/IEC 15408, хоча і випереджає НД ТЗІ, у контексті визначення ролі та необхідності криптографічної підтримки функцій захисту від НСД, але визначає лише вимоги до "Керування криптографічними ключами" (родина FCS_CKM) та вимоги з "Криптографічних операцій" (родина FCS_COP). З аналізу сфери застосування ISO/IEC 15408 випливає, що він не ставить за мету оцінити специфічні якості криптографічних алгоритмів і, отже, не містить у своєму складі відповідних критеріїв. Передбачається, що для підтвердження стійкості проводитиметься незалежна оцінка відповідних математичних властивостей.

У Канадських критеріях [3] питанню криптографічної підтримки присвячено додаток К, що визначає, яким чином криптографія може бути використана для реалізації послуг захисту від НСД. Користуючись тим, що НД ТЗІ є правонаступницею Канадських критеріїв визначимо, що надані рекомендації щодо криптографічної підтримки таких послуг як: довірча конфіденційність (КД-1, КД-2, КД-3), адміністративна конфіденційність (КА-1 та КА-2), повторне використання об'єктів (КО-1), ідентифікація та автентифікація (НИ-1, НИ-2, НИ-3), довірений канал (НК-1 та НК-2), довірча цілісність (ЦД-1, ЦД-2, ЦД-3), адміністративна цілісність (ЦА-1 та ЦА-2), розподіл обов'язків (НО-3).

Однак проведений аналіз не свідчить, що у ISO/IEC 15408 зовсім не висувається вимог до рівня стійкості послуг безпеки. Згідно ISO/IEC 15408-1 визначається, що стійкість функції (послуги) безпеки – характеристика функції безпеки, що відображає мінімальні зусилля, які необхідні для порушення політики безпеки цієї послуги, у разі здійснення атаки безпосередньо на механізми захисту, на яких вона ґрунтується.

У ISO/IEC 15408-1 визначено такі рівні стійкості як:

- базова стійкість – захист від випадкового порушення політики послуги безпеки з боку порушників із низьким потенціалом нападу;
- середня стійкість – захист від безпосереднього/навмисного порушення політики послуги безпеки з боку порушників із помірним потенціалом нападу;
- висока стійкість – захист від ретельно спланованого та організованого порушення політики послуги безпеки з боку порушників із високим потенціалом нападу.

Слід зазначити, що у серії стандартів ISO/IEC 15408 розгляд питання стійкості криптографічних алгоритмів не входить до сфери застосування, розглядаються лише імовірнісні та перестановочні методи, що не є криптографічними. З метою оцінювання рівня довіри до реалізації послуг безпеки спільним технічним комітетом ISO/IEC JTC 1, і його підкомітетом SC 27 був підготовлений стандарт ISO/IEC 18045 «Загальна методологія оцінки безпеки інформаційних технологій», що має на меті визначення засад оцінювання на рівнях довіри EAL1 – EAL4. Таким чином, для визначення стійкості послуг безпеки необхідно визначити потенціал порушника. У додатку А ISO/IEC 18045 [4] наведено методику оцінювання потенціалу порушника, що визначається на основі зусиль, що необхідні для ідентифікації та використання вразливостей. Множина факторів, що використовується при оцінюванні рівня потенціалу та відповідні оцінки зведені до табл. 4.

Із застосуванням критеріїв з табл. 4 експерт обирає значення, що характеризують здатність зловмисника ідентифікувати чи використати вразливість. Усі обрані значення додаються і на основі результуючої суми за табл. 5 визначаються рівень потенціалу зловмисника, а також необхідний рівень стійкості послуги безпеки.

Таблиця 4

Назва фактору	Діапазон	Значення для ідентифікації вразливості	Значення для використання вразливості
Час, що витрачається порушником	< 0,5 години	0	0
	< 1 доби	2	3
	< 1 місяця	3	5
	> 1 місяця	5	8
	Непрактично	*	*
Компетентність порушника	Непрофесіонал	0	0
	Професіонал	2	2
	Експерт	5	4
Знання об'єкту атаки	Відсутність інформації	0	0
	Загальнодоступна інформація	2	2
	Чутлива інформація	5	4
Час доступу до об'єкту атаки	< 0,5 години або доступ, що не може бути викрито	0	0
	< 1 доби	2	4
	< 1 місяця	3	6
	> 1 місяця	4	9
	Непрактично	*	*
Обладнання, що може використовуватися	Відсутнє	0	0
	Стандартне	1	2
	Спеціалізоване	3	4
	Розроблене на замовлення	5	6
* – Позначає, що напад неможливий у межах часових рамок, які були б прийнятні для порушника, отже будь-яке значення «*» вказує на високий рейтинг			

Таблиця 5

Діапазон значень	Послуга безпеки протистоїть порушнику з потенціалом нападу	Рівень стійкості послуги безпеки
< 10	Рейтинг відсутній	
10 – 17	Низький	Базовий
18 – 24	Помірний	Середній
> 24	Високий	Високий

Методи і моделі криптографічної підтримки послуг захисту від НСД та рівень їх стійкості

На сьогодні не існує загальноприйнятих підходів до криптографічної підтримки послуг захисту від НСД. Проте аналіз окремих рішень та публікацій, що були використанні при вирішенні завдань захисту від НСД та їх співставлення із специфікаціями, визначених у НД ТЗІ, дозволяє описати окремі моделі та методи забезпечення криптографічної підтримки.

Модель криптографічної підтримки адміністративної конфіденційності полягає у поділі множини суб'єктів на домени та призначенні кожному з доменів власного симетричного ключа шифрування. При цьому ключ шифрування має зберігатися тільки у КЗЗ, оскільки надання ключа шифрування користувачу дозволить останньому керувати потоками інформації. Такий підхід дозволяє вирішити задачу розмежування доступу користувачів до інформації, критичної з точки зору конфіденційності, як під час збереження, так і під час експорту за межі домену, що контролюється КЗЗ, за рахунок керування ключами шифрування.

Модель, що запропонована для реалізації послуги адміністративної конфіденційності, не може бути у явному вигляді застосована для криптографічної підтримки послуги довірчої конфіденційності. Найбільш доречним тут є використання методів асиметричної криптографії, які дозволять захищати за рахунок криптографічних функцій цілісність атрибутів доступу до об'єктів захисту. Зазначена властивість досягається за рахунок надання вповноваженому користувачу ключа створення/редагування списків контролю доступу.

Криптографічна підтримка послуг адміністративної цілісності та довірчої цілісності може забезпечуватися за рахунок криптографічного протоколу, що реалізується КЗЗ, та полягає у перевірці володіння вповноваженими користувачами атрибутами доступу – відкритим ключем з визначеної адміністратором ключової пари.

Послуга повторного використання об'єктів може бути підсилена за рахунок криптографічного протоколу шифрування вмісту спільно використовуваних об'єктів. Таке шифрування не дозволить не вповноваженим суб'єктам отримати доступ до звільнених іншим процесом/користувачем пасивних об'єктів.

Модель забезпечення послуги розподілу обов'язків припускає, що функції користувачів, які притаманні певним ролям, можуть бути активовані у разі надсилання керуючих команд, що мають бути перетворені із використанням особистого ключа, що є атрибутом ролі.

Згідно із специфікацією послуги достовірного каналу, КЗЗ має забезпечувати захист від несанкціонованої модифікації та ознайомлення з атрибутами доступу користувачів, що передаються на етапах автентифікації. Зазначена задача має вирішуватися за рахунок використання криптографічних протоколів шифрування та ЕЦП.

Криптографічні протоколи, що є функціями криптографічної підтримки послуг захищеного обміну даними (конфіденційність та цілісність при обміні), ідентифікації та автентифікації, а також неспростовності відправника і одержувача, є добре дослідженими і викладені у ряді міжнародних та національних стандартів [5, 6].

Модель криптографічної підтримки послуги контролю цілісності КЗЗ передбачає наявність у КЗЗ симетричних ключів (підтримання окремого домену), а також асиметричних ключів, що можуть бути використані для контролю цілісності власних складових, наприклад за рахунок застосування механізму електронного цифрового підпису.

Слід зазначити, що незважаючи на те, що криптографічні примітиви, які використовуються як функції криптографічної підтримки, мають складність яку можна оцінити кількісно, у рамках розглянутих моделей не може бути забезпечено доказового рівня стійкості послуг захисту від НСД. Насамперед, це пов'язано з обробкою у системі таємного компоненту, який може стати відомий за рахунок неправильної реалізації, спостереженням за побічними каналами, чи внаслідок організаційних помилок. І отже стійкість цих послуг безпеки може бути оцінено лише кількісно за методикою аналогічною до викладеної у додатку А ISO/IEC 18045.

Пропонується звести задачу з подолання механізмів захисту до стійкості функцій криптографічної підтримки послуг безпеки. Відмітимо, що такий підхід не суперечить поглядам закріпленим у серії стандартів ISO/IEC 15408 [2], оскільки відповідальність за оцінку стійкості використовуваних криптографічних функцій буде віднесено до сфери відповідальності органів сертифікації у сфері КЗІ. У роботі [7] запропоновано підхід до забезпечення доказового рівня стійкості послуг захисту від НСД.

Визначення 1. Функцією криптографічної підтримки послуги захисту від НСД називають деяку функцію, яка за рахунок використання методів КЗІ, ускладнює задачу зловмисника з подолання механізмів захисту, що реалізують послугу.

Визначення 2. Захищеним станом (SEC) називається стан, у якому із заданою імовірністю у зловмисника немає доступу до інформації HI, що дозволило б йому порушити правила політики безпеки, без подолання КЗЗ.

Визначення 3. КЗЗ забезпечує доказовий рівень стійкості механізмів захисту від загроз НСД, якщо складність отримання зловмисником інформації HI не менше від припустимого значення кількісного показника.

Визначення 4. Доказовий рівень стійкості КЗЗ від загроз НСД забезпечується із застосуванням функції криптографічної підтримки, якщо HI не використовується поза її межами, а також виконується хоча б одна з таких умов:

– значення показника стійкості КЗЗ від НСД є не меншим від значення показника стійкості функції криптографічної підтримки до відомих атак (умова $\Re 1$);

– HI є закритим (таємним) параметром функції криптографічної підтримки, що не обробляється у прикладній системі (умова $\Re 2$).

Слід зазначити, що у разі забезпечення доказового рівня стійкості, атака на КЗЗ не може бути здійснена за рахунок його помилкової (хибної) реалізації чи спостереження за залишковою інформацією. Отже фактори потенціалу нападу з ISO/IEC 18045: рівень компетентності порушника, час, що витрачається, знання об'єкту атаки, час доступу до об'єкту атаки, обладнання, що використовується, буде стосуватися виключно розв'язання задачі, яка лежить в основі стійкості криптографічного перетворення, що використовується як функція криптографічної підтримки.

У роботах [8 – 9] запропоновані методи захисту від НСД в частині забезпечення захисту від загроз порушення приватності. Основою цих методів є використання у якості функцій криптографічної підтримки: групового підпису та хамелеон-підпису у групі точок еліптичної кривою. Особливістю зазначених криптопримітивів є можливість побудови моделей політики безпеки, що дозволяють на практиці виконати умову $\Re 2$ та забезпечити доказовий рівень стійкості від загроз порушення приватності.

Висновки

1. Нормативна база України не враховує світових тенденцій в напрямку забезпечення приватності користувачів інформаційно-телекомунікаційних систем. Сформульовані у роботі рекомендації щодо специфікацій послуг приватності дозволяють розширити перелік функцій захисту, що регламентується НД ТЗІ 2.5-004-99.

2. Забезпечення захисту від загроз порушення приватності є комплексною задачею, що передбачає обґрунтування вибору та впровадження послуг анонімності, псевдонімності, неможливості асоціації та прихованості. Зазначені послуги є попарно незалежними і мають обиратися згідно із розробленою для конкретної ІТС моделі загроз.

3. У НД ТЗІ 2.5-004-99 відсутні критерії стійкості послуг безпеки. Рівні стійкості послуг безпеки з ISO/IEC 15408 та методологія оцінювання з ISO/IEC 18045 дозволяють оцінити стійкість послуг безпеки лише із застосування експертних оцінок на якісному рівні.

4. Використовувані сьогодні моделі та криптографічні протоколи підтримки "базових" послуг безпеки, пов'язаних із розмежуванням доступу, передачею каналами зв'язку, ідентифікацією/автентифікацією не можуть забезпечити доказовий рівень стійкості, оскільки передбачають обробку у системі таємного компоненту, що може бути скомпрометований, наприклад за рахунок побічних каналів витоку інформації.

5. Застосування механізмів групового підпису та методів хамелеон-гешування, хамелеон-підписів дозволяє надавати криптографічну підтримку послуг приватності із доказовим рівнем стійкості.

Список літератури: 1. *НД ТЗІ* 2.5-004-99. Критерії оцінки захищеності інформації в комп'ютерних системах від несанкціонованого доступу. 2. *ISO/IEC 15408* – 2. Information technology–Security techniques–Evaluation criteria for IT security – Part 2. Security functional requirements. 3. *Canadian* Security Establishment (CSE), "The Canadian Trusted Computer Product Evaluation Criteria (CTCPEC)", Version 3.0, January 1993. 4. *ISO/IEC18045*. Information technology – Security techniques – Methodology for IT security evaluation. 5. *ДСТУ* ISO/IEC 13888-1:1997 "Інформаційні технології. Методи захисту. Неспростовність. Ч. 1. Загальні положення". 6. *ДСТУ* ISO/IEC 9798-3:2006 "Інформаційні технології. Методи захисту. Автентифікація об'єктів. ". Ч. 3. Механізми, що ґрунтуються на використанні алгоритмів цифрового підпису. 7. *Лєншина, Ю.М.* Модель системи обслуговування замовлень з доказовим рівнем стійкості механізмів захисту від загроз приватності та неспростовності джерела / Ю.М. Лєншина // Системи обробки інформації. – Х., 2011. – Вип. 3(19). – С. 213 – 225. 8. *Погребняк, К.А.* Анализ современных групповых подписей на основе парных отображений и перспективы их использования в национальном электронном документообороте / К.А. Погребняк, Ю.М. Іщенко // Радіоелектронні і комп'ютерні системи. – Х., 2010. – Вип. 6(47). – С. 94 – 98. 9. *Лєншина, Ю.М.* Реалізація послуги приватності у схемі інтернет-аукціону з використанням хамелеон-підпису / Ю.М. Лєншина, Д.С. Беляк // Системи обробки інформації. – Х., 2011. – Вип. 3(93). – С. 126 – 129.

Харківський національний університет радіоелектроніки

Надійшла до редколегії 23.09.2012

УДК 681.03.06

К.А. ПОГРЕБНЯК, канд. техн. наук, Д.В. ПОВТАРЄВ

МОДЕЛЬ ВИКОРИСТАННЯ ХМАРНИХ ОБЧИСЛЕНЬ ДЛЯ ЗАДАЧ АСИМЕТРИЧНОГО КРИПТОАНАЛІЗУ НА ПРИКЛАДІ ФАКТОРИЗАЦІЇ ЧИСЕЛ МЕТОДОМ ρ-ПОЛЛАРДА

Вступ

Інформаційні технології традиційно є галуззю знань, в якій нові підходи розвиваються напрочуд швидко. Сьогодні одним з таких підходів є впровадження технології хмарних обчислень. Наразі створюється велика кількість нових хмарних сервісів, проходить етап міграції існуючих веб додатків до хмар, а також виходять публікації на тему використання хмарних рішень для все більшого класу задач. Все це свідчить про не аби яке зацікавлення технологією світовою спільнотою, зокрема провідними корпораціями та науковим співтовариством [1].

По-суті, хмарні обчислення є динамічно масштабованим способом доступу до зовнішніх обчислювальних ресурсів у вигляді сервісу, що надається за допомогою Інтернету, при цьому користувачам не потрібні особливі знання про інфраструктуру хмари або навички управління цією технологією. Фактично, використовуючи різні рівні хмарних обчислень, користувач буде отримувати доступ до деякого програмного забезпечення (рівень *SaaS – Software as a Service*), інтегрованої платформи (рівень *PaaS – Platform as a Service*) чи навіть комп'ютерної інфраструктури (рівень *IaaS – Infrastructure as a Service*), які будуть надаватися на вимогу як сервіси [2].

Слід відзначити, що на даному етапі потенціал хмарних обчислень не враховується в повному обсязі відносно їх впливу на інші технології, зокрема у галузі використання асиметричної криптографії.

За останні десятиріччя асиметрична криптографія поширилась на усі сфери життя сучасної людини. Її використання для економічних, політичних, військових задач визначається вимогою дотримання найвищого рівня безпеки, що ґрунтується на стійкості криптографічних примітивів. Оцінка стійкості криптографічних примітивів традиційно базується на виборі найбільш ефективного методу криптоаналізу та подальшому отриманні емпіричних характеристик обраного методу. Критерієм ефективності методу, як правило, вважається мінімізація обчислювальних ресурсів, тобто просторово часових показників. В той же час оцінки методів криптоаналізу у повному обсязі не враховують використання хмарних обчислень.

Таким чином, актуальною задачею є розробка моделі використання обчислювальних ресурсів у хмарі для різних методів криптоаналізу.

1. Модель використання хмарних обчислень для задач криптоаналізу

Застосування технології хмарних обчислень для задач криптоаналізу має ряд суттєвих переваг і, водночас, не позбавлено певних недоліків. Сукупність таких переваг та недоліків визначає вимоги до моделі використання хмарних обчислень у задачах криптоаналізу. Розглянемо більш детально переваги та недоліки використання хмарних обчислень та сформуємо основні вимоги до криптоаналітичної моделі.

1.1. Переваги та недоліки використання хмарних обчислень у задачах криптоналізу

Серед переваг технології хмарних обчислень, що мають суттєве значення для задач криптоаналізу, слід в першу чергу виділити наступні [3, 4]:

• Динамічна масштабованість системи. Обчислювальна потужність, яка доступна користувачу хмарних систем, практично обмежена лише розміром хмари, тобто загальною кількістю віддалених серверів. Користувачі можуть виконувати більш складні задачі, що потребу-

ють потужні процесори, велику кількість пам'яті та місце для зберігання даних, а система сама буде виділяти на це ресурси оптимальним чином та у необхідному обсязі.

• Спрощення адміністрування системи. Програмне забезпечення, що розгорнуто в хмарі, є захищеним та завжди підтримується в актуальному стані відповідними механізмами технології хмарних обчислень.

• Незалежність функціонування системи від пристроїв доступу до неї. Користувачі криптоаналітичної системи, побудованої на концепції хмарних обчислень, мають можливість доступу до неї фактично з будь-якого пристрою, який є підключеним до мережі Інтернет.

• Стійкість до непередбачених збоїв та фізична захищеність. Особливості функціонування хмарних систем забезпечують надійне виконання поставлених задач з низькою мірою залежності від стану певного апаратного вузла на якому виконуються обчислення.

Серед недоліків використання технології хмарних обчислень для задач криптоаналізу слід виділити наступні:

• Зберігання та контроль інформації за межами організації. Інформація, що зберігається за межами організації потенційно, може бути розкрита власниками системи хмарних обчислень, в якій вона циркулює. Власник криптоаналітичної системи не може в повному обсязі контролювати інформацію, що знаходиться в хмарі.

• Відсутність типових рішень та стандартів у зв'язку з новизною технології. Стандарти та технічні вимоги, що регулюють базові аспекти побудови хмарних систем та можуть в певному обсязі гарантувати безпеку інформації, знаходяться на етапі формування.

• Через особливості архітектури втрата даних в хмарі є безоборотною.

• Особливості створення програмних продуктів для розгортання у хмарах на різних платформах та, як наслідок, відповідні обмеження на міграцію від одного постачальника послуг до іншого.

1.2. Вимоги до моделі використання хмарних обчислень для задач криптоаналізу

Виходячи з переваг і недоліків технології хмарних обчислень можна висунути наступні вимоги до криптоаналітичної системи:

• Архітектура криптоаналітичної системи повинна бути багатомодульною. Модуль керування повинен бути відокремленим та знаходитись на комп'ютері, яким володіє користувач.

• Необхідність забезпечення моніторингу роботи криптоаналітичної системи та резервне копіювання цінної інформації.

• Незалежність криптоаналітичної системи від вибору платформи та забезпечення її функціонування при переході до будь-якого постачальника сервісу хмарних обчислень.

• Так як однією з найсуттєвіших причин використання хмарних технологій для задач криптоаналізу є низька вартість обчислювальних потужностей, то необхідно обирати найбільш економічно доцільний сервіс. Обчислювальні потужності повинні бути суттєво дешевшими ніж при створенні аналогічною за потужністю системи.

1.3. Опис моделі використання хмарних обчислень для задач криптоаналізу

Виходячи зі сформованих вимог, загальні компоненти криптоаналітичної системи, повинні включати:

- модуль керування криптоаналітичною системою;
- модуль моніторингу криптоаналітичної системи;
- модулі, що використовують обчислювальні ресурси:
  - на основі глобальної мережі Інтернет;
  - на основі локальної обчислювальної мережі;
  - на основі сервісу хмарних обчислень.

Графічне представлення загальних компонентів розподіленої криптоаналітичної системи наведено на рис. 1.

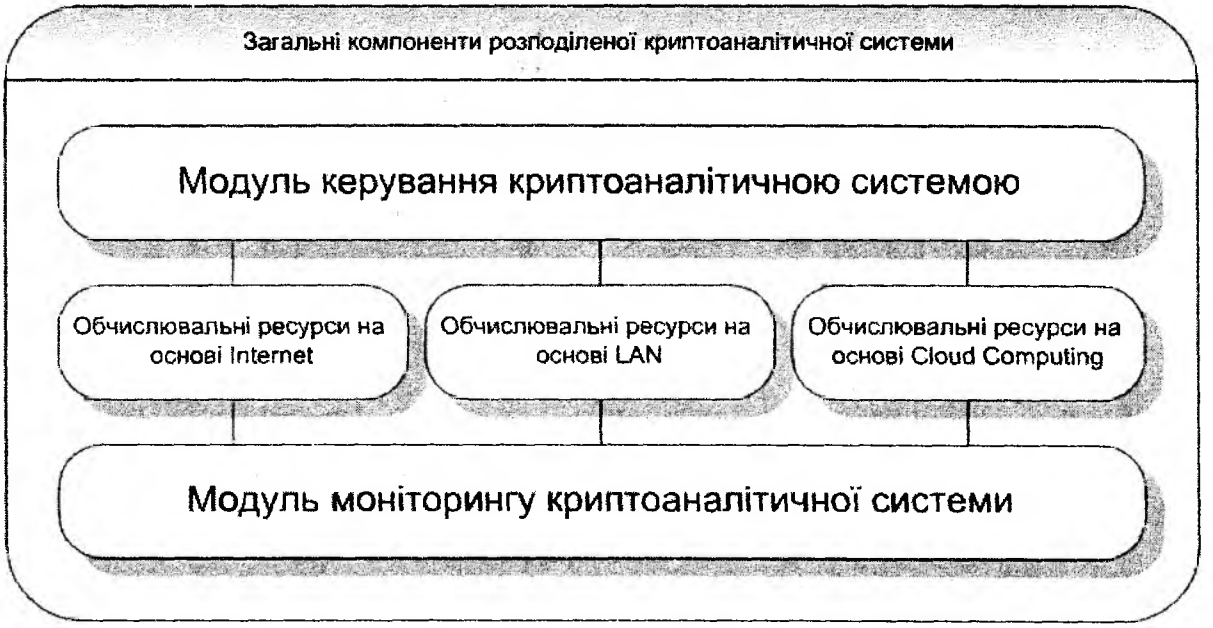


Рис.1

Для використання переваг технології хмарних обчислень та мінімізації ризиків, які пов'язані з нею, доцільним є одночасне використання модулів, що використовують усі обчислювальні ресурси.

Порівняння модулів, що використовують різні обчислювальні ресурси наведено в табл. 1.

Таблиця 1

Критерій порівняння	Обчислювальні ресурси на основі Internet	Обчислювальні ресурси на основі LAN	Обчислювальні ресурси на основі Cloud Computing
Стабільність роботи модуля системи	–	+	+
Контроль з боку власника системи	–	+	–
Наявність стабільного каналу передачі даних	+	–	+
Висока вартість володіння системою	–	+	–

Відмітимо, що основний модуль криптоаналітичної системи базується на використанні ресурсів локальної обчислювальної мережі. Такий модуль є найбільш надійним, оскільки він повністю контролюється власником системи, має фіксовану кількість робочих станцій, що гарантує певні показники потужності системи, та забезпечує контроль за резервним зберіганням інформації. Проте використання тільки ресурсів на основі локальної обчислювальної мережі є досить дорогим, так як вимагає коштів на створення та подальшу підтримку робочих станцій та серверів.

Модуль, що використовує обчислювальні ресурси у мережі Інтернет, є найменш стабільним та не забезпечує гарантійний рівень показників потужностей. З іншого боку такий компонент системи є абсолютно безкоштовним.

Модуль, що використовує ресурси на основі хмарних обчислень, є досить стабільним з гарантованим рівнем показників потужностей та зменшеною вартістю володіння системою в порівнянні з модулем, що ґрунтується на використанні ресурсів локальної обчислювальної

мережі. Недоліком такого компонента системи є неповний контроль з боку власника криптоаналітичної системи та наявність стабільного каналу передачі даних.

Використання різних обчислювальних ресурсів дозволяє визначати необхідний рівень потужності системи та вартість володіння нею.

2. Модель використання хмарних обчислень для задачі факторизації чисел методом ρ-Полларда

Безпека більшості сучасних криптосистем базується на складності задачі факторизації чисел.

Розглянемо побудування компонента системи, який використовує ресурси на основі хмарних обчислень, на прикладі розв'язання задачі факторизації чисел методом ρ-Полларда.

2.1. Опис методу ρ-Полларда факторизації чисел

Простота реалізації та відсутність залежності від структури групи зробили саме цей алгоритм найбільш прийнятним на цьому етапі досліджень. Метод ρ-Полларда базується на тому, що число N має такий простий дільник p, для якого $p-1$ є добутком невеликих простих чисел. Тому цей метод є ефективним при факторизації складених чисел з невеликими множниками у розкладенні.

Наведемо формальний опис алгоритму ρ-Полларда [5]:

Нехай N складене ціле додатне число.

Крок 1. Обираємо число k, що є добутком невеликих простих чисел в невеликих степенях. Наприклад, $k = НСК\{2,3,...M\}$ для деякого цілого додатного числа M.

Крок 2. Обираємо довільне число a, що задовольняє умові $1 < a < n$.

Крок 3. Обчислюємо найбільший спільний дільник $НСД(a,n)$. Якщо $НСД(a,n) > 1$ то ми отримуємо нетривіальний співмножник n. Якщо $НСД(a,n) = 1$, то переходимо до наступного кроку.

Крок 4. Обчислюємо $D = НСД(a^k - 1, n)$.

Якщо $1 < D < n$, то D є дільником числа N. Якщо $D = 1$, то повертаємося до кроку 1 та обираємо більший показник k. Якщо $D = n$ повертаємося до кроку 2 і обираємо нове число a.

2.2. Опис моделі використання хмарних обчислень для задачі факторизації чисел

Виходячи з наявної моделі та поставленої задачі побудуємо спрощену модель криптоаналітичної системи, яка буде використовувати ресурси тільки на основі хмарних обчислень.

Визначимо основні компоненти моделі:

- обчислювальний модуль криптоаналітичної системи;
- модуль керування криптоаналітичною системою;
- модуль моніторингу критоаналітичної системи;
- віддалений модуль керування криптоаналітичною системою;
- віддалений модуль моніторингу критоаналітичної системи.

Основними вимогами до спрощеної моделі є:

- наявність надійного канал Інтернет з'єднання;
- використання хмарного сервісу рівня PaaS;
- наявність засобу доступу до мережі Інтернет.

Графічний опис спрощеної моделі наведено на рис. 2.

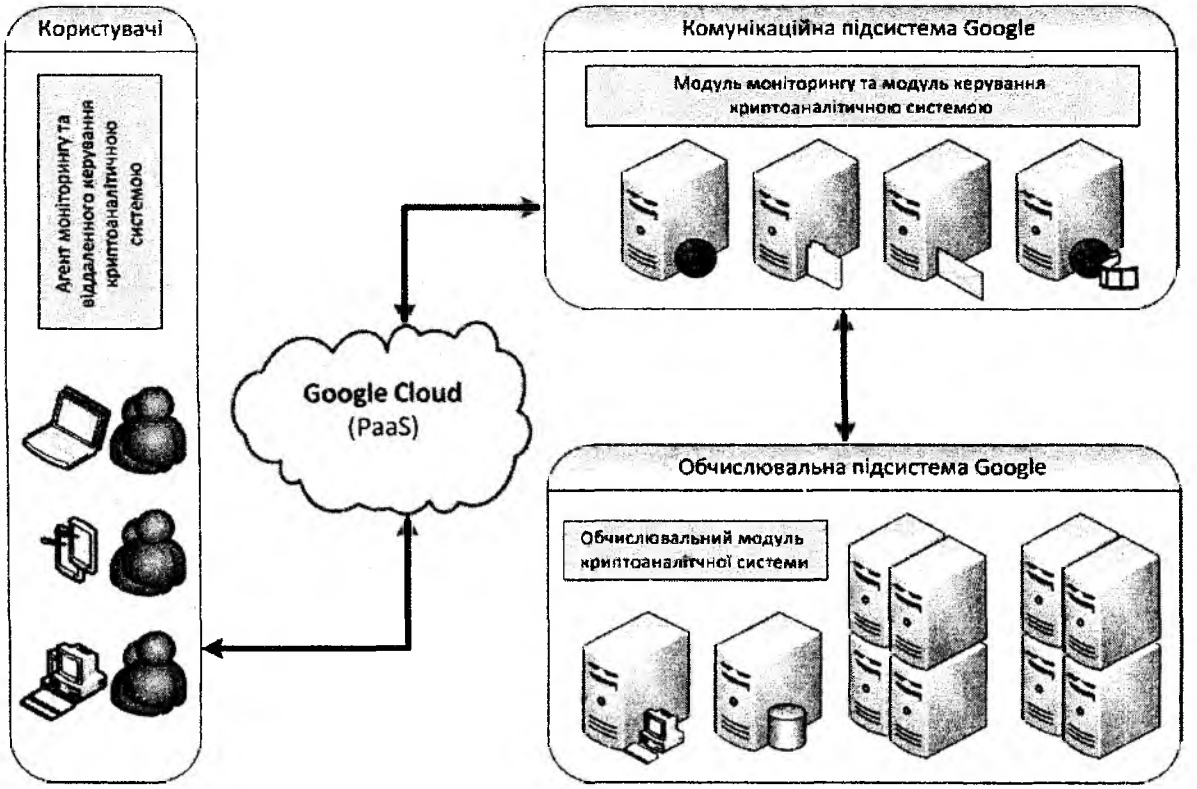


Рис.2

3. Реалізація моделі використання хмарних обчислень для задачі факторизації чисел методом ρ-Полларда з використанням технології Google App Engine

Для спрощення реалізації моделі використання хмарних обчислень для задачі факторизації чисел методом ρ-Полларда була обрана технологія Google App Engine (GAE), оскільки вона дозволяє розробляти, адмініструвати та масштабувати додатки, не зважаючи на апаратні засоби та резервне копіювання інформації. GAE надає доступ до інтегрованої платформи, тобто функціонує на рівні PaaS [6].

З застосуванням технології Google App Engine та мови програмування Java створено програмний продукт, що проводить факторизацію чисел за допомогою алгоритму ρ-Полларда на стороні сервера, тобто в хмарі Google [7].

Для підтвердження ефективності використання технологій хмарних обчислень для задач криптоаналізу проведене порівняння ефективності виконання факторизації за методом ρ-Полларда з використанням одного і того ж додатку, розгорнутому в хмарі та на персональному комп'ютері. Виконана факторизація чисел різної бітової довжини з різною кількістю співмножників.

Моделювання роботи програми проведено у такий спосіб. Проведено по 20 вимірювань для кожного довільно фіксованого складеного числа, мінімальні значення яких занесені до таблиці. Обчислення, що виконуються на стороні сервера, використовують потужності хмари і в кожен момент часу отримують необхідні ресурси, розмір яких контролює система. Проаналізувавши результати та порівнявши значення, можна зробити висновок, що зі збільшенням бітової довжини та кількості співмножників ефективність використання хмарних обчислень для задач криптоаналізу зростає.

Моделювання проводилось на комп'ютері з ОС – Microsoft Windows 7 Enterprise SP1 x86, процесором – Intel(R) Core (TM) 2 Duo 2,33 GHz, ROM – 3,50 Gb.

Результати виконання обчислень представлені у табл. 2.

Таблиця 2

№	Бітова довжина, *біт*	Кількість співмножників	Число/ Розклад числа	Час розкладу в хмарі, *мс*	Час розкладу на ПК, *мс*
1	74	5	**104353695894386059276 49** [313, 101, 745673, 254537, 1739173]	3	7
2	105	5	**353534543655432351365543457345 35** [5, 2803, 199, 1105774941389, 11463544797179]	4380	8845
3	105	19	**401553784227617377474369397350 40** [2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 7, 5, 23, 239, 123017, 11267303, 36762487387]	14	17
4	106	7	**493771156082948827115616533485 32** [2, 2, 3, 3, 3, 284927963, 1604600370135027022133]	182	192
5	106	13	**809873657435071567269852918946 20** [2, 2, 5, 11, 13, 101, 101, 107, 103, 211, 1739173, 745673, 907223]	3	16
6	107	13	**985287111111110898378666659948 80** [2, 2, 2, 2, 2, 2, 2, 5, 7, 47, 67, 35026304461, 199396627523279]	438	1867
7	107	19	**865123599856973844742274679582 72** [2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 107, 101, 1879, 4027, 254537, 907223, 745673]	5	7
8	107	19	**946175999999999795711999993548 80** [2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 7, 5, 7, 3, 35026304461, 199396627523279]	1017	1490
9	113	5	**958745583769558766685874558769558 7** [3, 3, 317, 1739173, 193222999807667261092252 3]	4	22
10	124	2	**109545874558376955876668587455876955 87** [6777349, 16163528624300881639217426674 63]	10	32
11	137	5	**109548834587455837695587666858745587695587** [3, 7, 17041, 139091, 22008698503808543904256456446 37]	3	9
12	148	4	**235109548834887455837675587666859745587695587** [3, 3, 379, 689268686118110395302479002248196263816 17]	2	3
13	153	4	**107154823645985214654897542313456564213246547 97** [23, 1852579, 1216358616623, 20674999945252002116276896 7]	10710	14154
14	166	4	**555672341095488345874558376955876668587455876955 87** [71, 431, 907223, 2001562285545371538453585678573842585069]	12	14
15	172	3	**567556723410954883458745583869558766685874558769558 7** [3, 60611333, 3121290443658750987810291648865725967374641 3]	140	325
16	182	7	**456545788200019878978978978562139835455677647987954340** [2, 2, 5, 331, 4027, 197161, 868607768752751109739690292551110374498981]	7	7
17	182	7	**456545788200019878978978978562139835455677647987954340** [2, 2, 5, 331, 4027, 197161, 868607768752751109739690292551110374498981]	9	10
18	184	5	**156755672341095488345874567750837695586668587455876955 87** [13, 17, 46021, 745673, 206693358512674889579069789337680023648885 9]	8	15
19	186	5	**756755669341995488345874586775083769558666858745587695 58** [2, 7, 13, 254537, 163355354357039542718497366231711249207334061613 7]	10	20
20	187	6	**985655332155653535554897946205026776206634866660489066 0** [2, 2, 3, 5, 105051229, 156376931775442841884868983719124738912012710 59]	100	207

Висновки

Робота присвячена вирішенню задачі розробки моделі використання обчислювальних ресурсів у хмарі для різних методів криптоаналізу.

Проаналізовано переваги та недоліки використання хмарних обчислень у задачах криптоналізу, на основі яких сформовано перелік вимог до криптоаналітичної системи.

Розроблено загальну модель використання хмарних обчислень для задач асиметричного криптоаналізу. Загальні компоненти криптоаналітичної системи повинні включати одночасно модулі, які використовують обчислювальні ресурси на основі глобальної мережі Інтернет, на основі локальної обчислювальної мережі та на основі сервісу хмарних обчислень. Але основний модуль криптоаналітичної системи повинен базуватися на використанні ресурсів локальної обчислювальної мережі, оскільки є найбільш надійним.

Визначено, що модуль, який використовує ресурси на основі хмарних обчислень, є досить стабільним та характеризується зменшеною вартістю володіння системою в порівнянні з модулем, що ґрунтується на використанні ресурсів локальної обчислювальної мережі. Тобто використання ресурсів на основі хмарних обчислень дозволяє зменшити вартість володіння в обмін на неповний контроль з боку власника криптоаналітичної системи та наявність стабільного каналу передачі даних.

На основі загальної моделі розроблена часткова модель для розв'язання задачі факторизації чисел методом ρ-Полларда, яка використовує ресурси тільки на основі хмарних обчислень. Реалізація часткової моделі використовує технологію Google App Engine та мову програмування Java.

За результатами моделювання процесу розв'язання задачі факторизації чисел методом ρ-Полларда у хмарі можна зробити висновок, що зі збільшенням бітової довжини та кількості співмножників ефективність використання хмарних обчислень для задач криптоаналізу зростає.

У подальшому планується розширити кількість методів криптоаналізу, відокремити модуль керування та реалізувати віддалений агент моніторингу.

Сервіс, що реалізує розв'язання задачі факторизації чисел методом ρ-Полларда, є загальнодоступним та знаходиться за адресою http://factorintegers.appspot.com.

Список літератури: 1. *John W. Rittinghouse, James F. Ransome.* Cloud Computing: Implementation, Management, and Security. CRC Press, 2009. 340 p. 2. *Tim Mather, Subra Kumaraswamy, Shahed Latif.* Cloud Security and Privacy: An Enterprise Perspective on Risks and Compliance. O'Reilly Media, 2009. 334p. 3. *Toby Velte, Anthony Velte, Robert Elsenpeter.* Cloud Computing, A Practical Approach. McGraw Hill Professional, 2009. 352 p. 4. *Barrie Sosinsky.* Cloud Computing Bible. John Wiley & Sons, 2010. 672p. 5. *Cormen, Thomas H.; Leiserson, Charles E.; Rivest, Ronald L. & Stein, Clifford.* Introduction to Algorithms (Second ed.). Cambridge, MA: MIT Press. 2001. 896–901pp. 6. *Sanderson, Dan.* Programming Google App Engine (1st ed.). O'Reilly Media. 2009. 400 p. 7. *Charles Severance.* Using Google App Engine. O'Reilly Media, Inc. 2009. 272 p.

Харківський національний університет радіоелектроніки

Надійшла до редколегії 23.09.2012

УДК 004.056.55

І.Д. ГОРБЕНКО, д-р техн. наук, Л.В. МАКУТОНІНА

АНАЛІЗ СТІЙКОСТІ ОБЧИСЛЮВАЛЬНИХ ЗАДАЧ, ЩО ЗАСНОВАНІ НА БІЛІНІЙНИХ ВІДОБРАЖЕННЯХ

Вступ

Одним із перспективних напрямів розвитку криптографічних систем на ідентифікаторах є поєднання математичного апарату бінарного відображення (спарювання) точок еліптичних кривих з алгебраїчними решітками [1]. Таке поєднання дозволяє використати переваги спарювання та алгебраїчних решіток, основними з яких є зменшена складність (підвищена швидкодія) обчислень при криптографічних перетвореннях, відмова від використання сертифікатів, порівняно невеликий розмір зашифрованих текстів та ключових даних. Але, як для окремо взятих, так для поєднаного криптографічного перетворення з використанням удосконаленого методу, в першу чергу потрібно порівняти їх за безумовним критерієм криптографічної стійкості та складності (швидкодії) криптографічних перетворень.

Метою цієї статі є аналіз основних обчислювальних задач криптографії з використанням білінійних відображень та, по можливості, визначення основних питань їх використання при поєднанні бінарного відображення точок еліптичних кривих з алгебраїчними решітками.

Основні положення білінійних відображень

Класичні схеми на ідентифікаторах використовують математику білінійних відображень (спарювань). В криптографічних схемах на ідентифікаторах використовується не вироджене, ефективне для обчислювання білінійне відображення:

$$e : G_1 \times G_2 \to G_3, \tag{1}$$

де G_1, G_2, G_3 – циклічні групи однакового порядку p. Найчастіше складовими відображення є елементи еліптичних кривих, таким чином, G_1 і G_2 виступають циклічними підгрупами точок на еліптичній кривій над кінцевим полем, а результатом відображення є підгрупа G_3 мультиплікативної групи над кінцевим полем. Зазвичай груповою операцією для еліптичних кривих є додавання, тому G_1 і G_2 є адитивними підгрупами, в той час, як G_3 є мультиплікативною підгрупою [1]. Основними властивостями білінійного відображення є невиродженість, ефективність обчислення та білінійність.

Невиродженість – це властивість, при якій якщо $e(P,Q)$ тотожно елементу G_3, тоді P тотожно G_1 та/або Q тотожно G_2. Тобто виконуються такі умови:

1) для всіх $P \in G_1$, з $P \neq 0$, існує деяке $Q \in G_2$, таке, що $e(P,Q) \neq 1$;

2) для всіх $Q \in G_2$, з $Q \neq 0$, існує деяке $P \in G_1$, таке, що $e(P,Q) \neq 1$.

Під ефективністю мається на увазі, що існує поліноміальний алгоритм, що дозволяє ефективно обчислювати відображення $e(X,Y)$.

Під білінійністю розуміється лінійність відображення e по обом компонентам, що задовольняють таким двом властивостям:

$$\begin{aligned} e(R_1 + R_2, Q) &= e(R_1, Q)e(R_2, Q); \\ e(R, Q_1 + Q_2) &= e(R, Q_1)e(R, Q_2). \end{aligned} \tag{2}$$

Наслідком наведених властивостей є наступне співвідношення, яке і є основною відмінністю всіх схем, що засновані на білінійних відображеннях. Сутність його в тому, що для $a,b \in Z_p$ справедливим є

$$e(aP_1, bP_2) = e(bP_1, aP_2) = e(P_1, P_2)^{ab}. \tag{3}$$

Важливими також є і такі наслідки білінійності.

Нехай $P \in G_1$ і $Q \in G_2$, тоді для білінійного спарювання e виконуються наступні властивості:

1) $e(P,0) = e(0,Q) = 1$, що є наслідком (2), так як $e(P,Q) = e(P+0,Q) = e(P,Q)e(0,Q)$;

2) $e(-P,Q) = e(P,Q)^{-1} = e(P,-Q)$, так як $1 = e(0,Q) = e(P+(-P),Q) = e(P,Q)e(-P,Q)$;

3) $e(jP,Q) = e(P,Q)^j = e(P,jQ)$, для всіх $j \in Z$.

Визначення дивізора

Ключовим поняттям для білінійних відображень є теорія дивізорів (дільників) алгебраїчної кривої, яку запропонував Андре Вейль. Спираючись на [2 – 4] дамо визначення дивізора. Нехай $E: y^2 = x^3 + ax + b$ – еліптична крива над ідеальним полем K, та нехай $E(\overline{K})$ позначає алгебраїчне замикання поля K, тобто, множину точок $P(x,y) \in EC(\overline{K})$, що задовольняють $nP = \infty$, таким чином, поле K – кінцеве, а його замикання нескінчене. Тоді, дивізором над кривою E називають формальну суму

$$D = \sum_{P \in E(\overline{K})} n_P [P], \tag{4}$$

де коефіцієнти $n_P \in Z$ – цілі числа, причому число доданків з ненульовими коефіцієнтами n_P є кінцевим. Множина дивізорів еліптичної кривої утворює адитивну групу відносно операції додавання, а нулем є дивізор, у якого всі коефіцієнти дорівнюють нулю, тобто дивізор з усіма $n_P = 0$ дорівнює нулю. Множина дивізорів на кривій E позначається як $div_{\overline{K}}(E)$, а сума дивізорів визначається природним шляхом. Множина точок P, для яких $n_P \neq 0$ зветься носієм (support) дивізора D і позначається $supp(D)$. Ціле число $k = \sum n_P$, $P \in supp(D)$, зветься ступенем D і позначається $deg(D)$.

Визначимо також поняття групи Галуа. Групою Галуа називають групу ізоморфізмів над певним розширенням поля. Нехай $\overline{K}$ є розширенням поля K. Нехай, визначена множина всіх автоморфізмів поля $\overline{K}/K$, тобто ізоморфізм α з поля $\overline{K}$ сам у себе такий, що $\alpha(x) = x$ для всіх $x \in K$, тоді розширення $\overline{K}/K$ називається розширенням Галуа, і, для нашого випадку, група Галуа позначається як $\sigma \in Gal(\overline{K}/K)$. Визначимо $D^\sigma = \sum_P n_P(\sigma(P))$. Дивізор визначається над K, якщо $D = D^\sigma$ для всіх $\sigma \in Gal(\overline{K}/K)$.

Нехай $f(x,y): E \to K$ – раціональна функція. Якщо f не є константою, тоді існує кінцеве число точок $P \in E$, в яких $f(P) = 0$ або $f(P) = \infty$. Точки виду $f(P) = 0$ називаються нулями функції f, точки виду $f(P) = \infty$ називаються полюсами функції f. З точністю до ненульового множника можна задати функцію f, перераховуючи всі її нулі і полюса, задаючи їх кратність. Якщо f має нуль/полюс кратності k в точці P, тоді функцію f можна представити у вигляді добутку $f = u_P^k \cdot g$, де u_P має нуль/полюс першого порядку в точці P, а $g(P) \neq 0$ і $g(P) \neq \infty$, відповідно. Функція u_P називається уніформі затором функції f в точці P. Якщо f – ненульова функція над кривою E, тоді $ord_P(f)$ містить безліч f для P. Необхідно відмітити, що $ord_P(f)$ є позитивним, коли $f(P) = 0$, і негативним, якщо f має полюс в точці P. Дивізор ненульової функції f, записаної як (f), є дивізор $\sum_{P \in C(\overline{K})} ord_P(f)(P)$. Звідси випливає, що $(fg) = (f) + (g)$ і $(f/g) = (f) - (g)$.

В групі дивізорів найважливішими є, так звані, головні дивізори (principaldivisors). Головний дивізор над кривою E – це дивізор, що дорівнює (f), для деякої функції f, причому $\deg((f)) = 0$ [див. 2,3]. Функція f визначається над K, якщо вона може бути записана

з усіма коефіцієнтами в полі K. Якщо f ненульова функція, визначена над K, тоді дивізор (f) визначено над K. Якщо $f \in \overline{K}^*$ – константа, тоді $(f)=0$, і, навпаки, якщо $(f)=0$, тоді f повинна бути константою. Іншими словами, дивізор (f) визначає функцію f з точністю до ненульового множника.

Два дивізора D і D' є еквівалентними $D \sim D'$, якщо $D' = D + (f)$ для деякої функції f. Якщо $D_1 \sim D'_1$ і $D_2 \sim D'_2$, тоді $(D_1 + D_2) \sim (D'_1 + D'_2)$. Групою класів дивізорів над кривої називають множину всіх дивізорів нульового ступеня, з груповою структурою, що еквівалентна $div_{\overline{K}}(E)$.

Визначимо також поняття групового гомоморфізму з адитивної групи дивізорів до мультиплікативної групи поля K. Нехай $P_1(x_1, y_1)$ і $P_2(x_2, y_2)$ – дві точки на кривій $E(K)$. Припустимо, що пряма, проведена повз точки P_1 і P_2 має вигляд $l(x,y)=0$. Якщо l не є дотичною в точках P_1 і P_2, тоді існує деяка третя точка $S(x_s, y_s)$, в який вона перетинає E. Лінійний многочлен $l(x,y)$ можна інтерпретувати, як функцію відображення точки еліптичної кривої в K. Функція $l = ax + by + c$ має дивізор $div(l_{P_1,P_2}) = 1[P_1] + 1[P_2] + 1[S] - 3[O]$. Вертикальна пряма $v(x) = (x - x_s)$, що проходить через точки S і $P_3 = P_1 + P_2$, і є функцією на кривій E, має дивізор $div(v_{S,P_3}) = 1[S] + 1[P_3] - 2[O]$. Таким чином, рівняння $P_3 = P_1 + P_2$ є еквівалентним рівнянню дивізора $[P_3] - [O] = [P_1] - [O] + [P_2] - [O] - div(l/v)$ і, тоді, відображення P до дивізору класу $[P] - [O]$ є груповим гомоморфізмом.

Визначимо поняття дивізора нульового ступеню. Нехай E – еліптична крива над полем K. Тоді, дивізор нульового ступеню на кривій E є дивізор, виду $D = \sum_P n_P(P)$. Далі, дивізор D з нульовим ступенем є дивізором деякої функції f тоді і тільки тоді, коли $\sum_P [n_P]P = O$, тобто тоді, коли сума дивізорів для цієї функції дорівнює нескінченності.

Нехай f – деяка функція, і нехай $D = \sum_P n_P(P)$ – дивізор нульового ступеню, такий, що $supp(D)$ не перетинається з $supp(f)$. Визначимо

$$f(D) = \prod_P f(P)^{n_P}. \tag{5}$$

Якщо $g = cf$, для деякої константи $c \in \overline{K}^*$, тоді, якщо D нульового ступеню, $g(D) = f(D)$. Отже, в цьому випадку величина $f(D)$ залежить тільки від дивізорів (f) і D. Якщо f і D визначені над полем K, тоді $f(D) \in K$.

З урахуванням наведеного вище можна дати визначення закону взаємності Вейля (Weil-reciprocity). *Нехай f і g будуть ненульовими функціями кривої C над полем K. Якщо $div(f)$ і $div(g)$ не мають спільних точок, тобто $supp(f)$ і $supp(g)$ не перетинаються, тоді $f(div(g)) = g(div(f))$.*

Спарювання Тейта та його основні властивості

В якості основних в криптографії використовується два типи спарювань – спарювання Тейта і спарювання Вейля [5]. Спочатку дамо визначення спарюванню Тейта. Існує декілька варіантів спарювання Тейта. Так, Тейт запропонував досить узагальнене визначення спарювання на абелевих різноманіттях над локальними полями. Пізніше, було запропоновано спарювання на кривій Якобіана над локальними полями, яке можна явно обчислити [6]. Фрей і Рак у роботах [7, 8] ввели визначення спарювання Тейта над кінцевим полем, яке зараз використовується в криптографії.

Нехай $E: y^2 = x^3 + ax + b$ – еліптична крива над полем K_0 та нехай n буде позитивним цілим, яке є взаємно простим з характеристикою поля K_0. Нехай також множина коренів n-го ступеня буде визначена як $\mu_n = \{u \in \overline{K}_0^* : u^n = 1\}$. Визначимо поле $K = K_0(\mu_n)$, розширення K_0 породжених n коренів з одиниці. Визначимо також:

$$E(K)[n] = \{P \in E(K) : [n]P = O\},$$
$$nE(K) = \{[n]P : P \in E(K)\}. \tag{6}$$

де $E(K)[n]$ – група показників ступеню n; $nE(K)$ – підгрупа $E(K)$, множина точок $\{nQ \mid Q \in E\}$;

Тоді, $E(K)/nE(K)$ – фактор-група ступенів n, множина класів еквівалентності кривої $E(K)$ по множині $nE(K)$, над відношенням $P_1 \equiv P_2$ еквівалентності, в тому випадку, якщо $(P_1 - P_2) \in nE(K)$. Також визначимо:

$$(K^*)^n = \{u^n : u \in K^*\}. \tag{7}$$

де $(K^*)^n$ – підгрупа групи K^* і фактор-група $K^*/(K^*)^n$ – група ступенів n. Групи $K^*/(K^*)^n$ і μ_n є ізоморфними.

Групи $E(K)[n]$ і $E(K)/nE(K)$ мають однакове число елементів, але точки групи $E(K)[n]$ можуть і не бути представниками класів групи $E(K)/nE(K)$. Так, наприклад, нехай p і r – прості числа, такі, що r^4 ділить $(p-1)$. Існує еліптична крива E над F_p з $p-1$ точками, що має r^4 точок порядку r^2 визначених над F_p. В даному випадку $E[r] \subseteq rE(F_p)$.

Нехай $P \in E(K)[n]$, $Q \in E(K)$. Вважаємо, що Q представляє клас еквівалентності в $E(K)/nE(K)$. Доки $[n]P = O$, існує функція f така, що $(f) = n(P) - n(O)$. Нехай D буде будь-яким дільником нульового ступеню еквівалентним $(Q) - (O)$, таким, що D визначено над K і основа D не перетинається з основою (f). У більшості випадків такий дільник може бути легко побудований шляхом вибору довільної точки $S \in E(K)$ і визначається $D = (Q + S) - (S)$. Оскільки f і D визначені над K, значення $f(D)$ є елементом K. Доки основи (f) і D не перетинаються, маємо $f(D) \neq 0$, і тоді $f(D) \in K^*$.

Спарювання Тейта над P і Q визначається як

$$\langle P, Q \rangle_n = f(D), \tag{8}$$

та інтерпретується як елемент $K^*/(K^*)^n$. Значення спарювання Тейта є класом еквівалентності, тобто символ «=» використовується для позначення еквівалентності над цим відношенням, доказ цього твердження наведено в [5].

Тобто, спарювання Тейта – це білінійне відображення виду

$$\tau_n : E[n] \times E/E[n] \to F_{q^k}^* \times F_{q^k}^* \setminus \mu_n, \tag{9}$$

де μ_n – мультиплікативна підгрупа коренів n-го ступеню з одиниці поля F_{q^k}, визначена як:

$$\tau_n(T, S) = \frac{f_T(S + R)}{f_T(R)}, \text{ де } R \notin \{T, -S, T - S, \infty\}. \tag{10}$$

Визначимо основні властивості спарювання Тейта. Нехай E – еліптична крива над K_0 і n – взаємно просте з характеристикою K_0. Нехай $K = K_0(\mu_n)$. Тоді спарювання Тейта задовольняє наступним властивостям:

1) Білінійність. Для всіх $P, P_1, P_2 \in E(K)[n]$ і $Q, Q_1, Q_2 \in E(K)/nE(K)$ справедливо:

$$\langle P_1 + P_2, Q\rangle_n = \langle P_1, Q\rangle_n \langle P_2, Q\rangle_n,$$
$$\langle P, Q_1 + Q_2\rangle_n = \langle P, Q_1\rangle_n \langle P, Q_2\rangle_n.$$

2) Невиродженість. Нехай K – кінцеве поле. Для всіх $P \in E(K)[n]$, $P \neq O$, існує деяке $Q \in E(K)/nE(K)$, таке, що $\langle P, Q\rangle_n \neq 1$.З іншого боку, для всіх $Q \in E(K)/nE(K)$ з $Q \notin nE(K)$ існує деяке $P \in E(K)[n]$, таке, що $\langle P, Q\rangle_n \neq 1$.

3) Інваріантність Галуа. Якщо $\sigma \in Gal(\overline{K}/K_0)$, тоді $\langle \sigma(P), \sigma(Q)\rangle_n = \sigma(\langle P, Q\rangle_n)$.

Доказ наведених властивостей наведено у джерелах [5, 7, 9]. Найважливішою є властивість невиродженості, на відміну від спарювання Вейля, відображення Тейта не дорівнює одиниці, при $P = Q$, ця властивість дозволяє обчислити деяке m, таке, що $Q = mP$ за одне обчислення. Відмітимо також, що значення перетворення Тейта $\tau(P, Q)$ визначається точками P і Q не однозначно, а з точністю до множника, що належить до групи μ_n. Для того щоб отримати конкретне унікальне значення, елемент $\tau(P, Q)$ підносять до ступеня $(q^k - 1)/n)$, тобто, $\tau_{un}(P, Q) = \tau(P, Q)^{(q^k-1)/n}$.

Спарювання Вейля

Нехай $E: y^2 = x^3 + ax + b$ – еліптична крива, визначена над алгебраїчно замкнутим полем точок кривої K_0 та нехай n буде позитивним цілим яке є взаємно простим з характеристикою поля K_0, і $E[n]$ – підгрупа точок кривої E порядку n: $E[n] = \{P \in E \mid n \cdot P = \infty\}$. Також нехай $K = K_0(E[n])$ буде розширенням поля K_0, що генероване координатами усіх точок поля $E(\overline{K})$, порядку, що ділиться на n. Спарювання Вейля – це відображення:

$$e_n : E[n] \times E[n] \to \mu_n \subseteq K^*. \tag{11}$$

Нехай $P, Q \in E[n]$ і $D_P = [P] - [O]$ і $D_Q = [Q] - [O]$ – два дивізори нульового ступеню, такі, що їх носії не перетинаються. Також нехай $f(P) = nD_P$ і $f(Q) = nD_Q$ будуть двома функціями, тоді спарювання Вейля можна визначити наступним чином:

$$e_n(P, Q) = \frac{f_P(D_Q)}{f_Q(D_P)}. \tag{12}$$

Спарювання Вейля приймає значення в $\mu_n \subseteq K^*$, де μ_n – це мультиплікативна підгрупа коренів ступеню n з одиниці поля K. Тоді спарювання Вейля задовольняє наступним властивостям:

1) Білінійність. Для всіх $P, P_1, Q, Q_1 \in E[n]$ справедливо:

$$\begin{aligned} e_n(P + P_1, Q) &= e_n(P, Q) e_n(P_1, Q), \\ e_n(P, Q + Q_1) &= e_n(P, Q) e_n(P, Q_1). \end{aligned} \tag{13}$$

2) Змінність. $e_n(P, P) = 1$, тоді $e_n(P, Q) = e_n(Q, P)^{-1}$.

3) Невиродженість. Якщо $e_n(P, Q) = 1$, для всіх $Q \in E[n]$, тоді $P = Q$.

4) Інваріантність Галуа. Для всіх $\sigma \in Gal(\overline{K}/K)$, справедливо

$$e_n(\sigma(P),\sigma(Q)) = = \sigma(e_n(P,Q)).$$

5) Сумісність. Якщо $P \in E[nm]$ і $Q \in E[n]$, тоді $e_{nm}(P,Q) = e_n([m]P,Q)$.

6) Якщо $\phi : E \to E'$ – ізоендоморфізм з подвійним $\hat{\phi}$, тоді $e_n(\phi(P),Q) = e_n(P,\hat{\phi}(Q))$.

Усі наведені властивості, окрім невиродженості, можна довести, використовуючи закон взаємності Вейля [див. 2].

Найважливішим наслідком наведених властивостей є наступне твердження. *Нехай E – еліптична крива над полем k, і, нехай, r – просте число. Нехай $P,Q \in E(k)[r]$, такі, що $P \neq Q$. Тоді Q лежить в підгрупі, що породжена P, тоді і тільки тоді, коли $e_r(P,Q) = 1$.*

Порівняльний аналіз спарювання Тейта і Вейла

Спочатку звернемо увагу на деяку схожість між визначеннями спарювання Тейта і спарювання Вейля. Для спарювання Вейля функція $f(D')$ еквівалентна по модулю n– го ступеню до $\langle P,Q\rangle_n$, тоді, як $g(D)$ є еквівалентним по модулю n– го ступеню до $\langle Q,P\rangle_n$. За цих умов можна записати:

$$e_n(P,Q) = \frac{\langle P,Q\rangle_n}{\langle Q,P\rangle_n}, \tag{14}$$

з точністю до n– го ступеня, якщо $\mu_n \subseteq (K^*)^n$.

Однією із відмінностей між спаровуваннями Вейля і Тейта, є той факт, що для спарювання Тейта використовується поле $K_0(\mu_n)$, а для спарювання Вейля потенційно набагато більше за розміром поле $K_0(E[n])$. У роботі [10] Кобліц та ін. показали, що у більшості випадках ці два поля співпадають. Дійсно, нехай E – еліптична крива над полем F_q, r – просте число, що ділить $\#E(F_q)$. Припустимо також, що r не ділить $(q-1)$ і НОД$(r,q) = 1$, тоді $E[r] \subset E(F_{q^k})$, тоді і тільки тоді, коли r ділить $(q^k - 1)$.

Спарювання Вейля може бути узагальнено з $E[n] = \ker([n])$ до $\ker(\phi)$, де ϕ – ізоендоморфізм. Так, у роботі [11] було показано, що для деяких випадків узагальнене спарювання Вейля по суті є еквівалентним спарюванню Тейта.

Аналіз стійкості основних обчислювальних задач, що засновані на білінійних відображеннях

Розглянемо основні задачі, наведені в літературі, що засновані на білінійних відображеннях, і є важко розв'язувальними. Стійкість наступних задач базується на твердженні, що на даний момент не існує ні єдиного алгоритму, який міг би вирішити дану задачу за поліноміальний час. Базовою задачею для криптосистем на ідентифікаторах, що використовують спарювання, є задача дискретного логарифмування [1, 5]. Для циклічної групи, що найчастіше використовується в криптографічних додатках, задача дискретного логарифмування у групі є складно обчислювальною, тобто не існує жодного алгоритму, який би вирішував цю задачу за поліноміальний час. Обчислювальна складність інших задач залежить від складності розв'язку задачі дискретного логарифму. Тому, спочатку дамо визначення задачі розв'язання дискретного логарифму над циклічною групою $\langle g\rangle$, але перед цим визначимо типи спарювань, що будуть необхідні в подальшому.

Нехай $e : G_1 \times G_2 \to G_T$ – білінійне відображення, що визначено над групами еліптичних кривих. Нехай G_1 і G_2 – циклічні групи однакового простого порядку, які є ізоморфними.

В залежності від структури групи G_2 білінійні спарювання можна класифікувати наступним чином [12]:

Тип 1. Спарювання належить до першого типу, якщо $G_1 = G_2$.

Тип 2. Спарювання належить до першого типу, якщо можна ефективно обчислити ізоморфізм ψ із G_2 до G_1.

Тип 3. Спарювання належить до першого типу, якщо не можна ефективно обчислити ізоморфізм із G_2 до G_1.

Для Типу 2 ізоморфізм є легко обчислювальним, тоді, як для Типу 3 на даний момент не існує даного ізоморфізму, але для двох даних типів припускається, що не існує ефективно обчислювального ізоморфізму з G_1 до G_2. Під Типом 1 розуміється використання симетричного спарювання, тобто $e(P,Q) = e(Q,P)$. Тип 2 і Тип 3 належать до асиметричного спарювання. Так як на практиці спарювання третього типу зазвичай не використовується, тому такі задачі не розглядатимуться в цій статті.

1) DLP-задача (Discrete Logarithm Problem).

Вхідні дані: Циклічна група $\langle g \rangle$, порядку p і елемент $h \in \langle g \rangle$.

Завдання: Обчислити $a \in Z_p$, що $h = g^a$.

Тобто, a зветься дискретним логарифмом від h, щодо основи g, і записується, як $\log_g h = a$. Для циклічної групи, що найчастіше використовується в криптографічних додатках, задача дискретного логарифмування у групі є складно обчислювальною, тобто не існує жодного алгоритму, який би вирішував цю задачу за поліноміальний час. Нехай також, основа логарифму над $\langle g \rangle$ має розмір n. Очевидно, можна знайти a шляхом повного перебору всіх елементів $\langle g \rangle$ за час $O(2^n)$, тобто за експоненціальний час. Звичайно, існують загальні алгоритми, такі, як, наприклад, метод Полларда, що дозволяють знизити складність до $O(2^{n/2})$, але вона все-таки залишається експоненціальною.Тобто, можна сказати, що на даний момент, задача дискретного логарифмування для деяких груп кінцевого поля та на еліптичних кривих є обчислювально складною і широко використовується у криптографії.

2) CDH-задача (Computational Diffie-Hellman).

Вхідні дані: Циклічна група $\langle g \rangle$, порядку p і елементи, де $a,b \in Z_p$ – рівномірні і випадкові елементи.

Завдання: Обчислити g^{ab}.

Іншими словами, якщо криптоаналітик (далі – КРА) A має мету вирішити CDH-задачу, за вхідні дані він має (g, g^a, g^b), а на виході повинен отримати g^{ab}. Тобто, A повинен вирішити наступне рівняння:

$$Adv^{CDH}(A) = \Pr[(g, g^a, g^b) \to g^{ab}]\,(g, g^a, g^b) \qquad (15)$$

Говорять, що CDH-задача є (ε, t)-складною, якщо для будь-якого A щонайбільше за час t, $Adv^{CDH}(A) \le \varepsilon$.

3) DDH-задача (Decisional Diffie-Hellman)

Вхідні дані: Циклічна група $\langle g \rangle$, порядку p і елементи (g, g^a, g^b, g^c), де $a,b \in Z_p$ – рівномірні і випадкові елементи.

Завдання: Визначити, чи є $c = ab$, або, чи $c \in Z_p$ випадковим рівномірним елементом.

Тобто, A за вхідні дані має (g, g^a, g^b, g^c), а на виході повинен отримати біт. Подія, при який A отримує одиницю, визначається, як $A(g, g^a, g^b, g^c) \to 1$. Тобто, A повинен вирішити наступне рівняння:

$$Adv^{DDH}(A) = |\Pr[A(g, g^a, g^b, g^c) \to 1] - \Pr[A(g, g^a, g^b, g^c) \to 1]|. \quad (16)$$

Говорять, що DDH-задачу в групі $\langle g \rangle$ є (ε, t)-складною, якщо для будь-якого A щонайбільше за час t, $Adv^{DDH}(A) \le \varepsilon$.

Далі дамо визначення декільком задачам для симетричного спарювання Типу 1, тоді, для даного типу $G = G_1 = G_2$, для позначення набору параметрів будемо використовувати (p, G, G, G_T, e). Нехай також, $G = \langle P \rangle$ і $G_T = \langle e(P, P) \rangle$. Також, припускається, що (p, G, G, G_T, e), а також генератори G і G_T та деякі інші параметри, в залежності від конкретної задачі, є відкритими параметрами.

Базовою складною задачею на білінійних відображеннях є білінійна задача Діффі – Гелмана, вперше запропонована Воне і Франкліним у [13].

4) BDH (BilinearDiffie-Hellman)

Вхідні дані: Набір (P, aP, bP, cP), де $a, b, c \in Z_p$ – рівномірні і випадкові елементи.

Завдання: Обчислити $e(P, P)^{abc}$.

Як і у випадку CDH-задачі, A повинен вирішити наступне рівняння:

$$Adv^{BDH}(A) = \Pr[A(P, aP, bP, cP) \to e(P, P)^{abc}] \quad (17)$$

Говорять, що DDH – задача у групі $\langle g \rangle$ є (ε, t)-складною, якщо для будь-якого A щонайбільше за час t, $Adv^{BDH}(A) \le \varepsilon$.

5) DBDH-задача (DecisionalBilinearDiffie-Hellman)

Вхідні дані: Набір (P, aP, bP, cP, Z), де $G = \langle P \rangle$, $a, b, c \in Z_p$ – рівномірні і випадкові елементи і $Z \in G_T$.

Завдання: Визначити, чи є $Z = e(P, P)^{abc}$, або, чи Z випадковим елементом G_T.

Тобто, A за вхідні дані має (P, aP, bP, cP, Z), а на виході повинен отримати біт. Подія, при який A отримує одиницю, визначається як $A(P, aP, bP, cP, Z) \to 1$. Тобто, A повинен вирішити наступне рівняння:

$$\begin{aligned} Adv^{DBDH}(A) = |&\Pr[A(P, aP, bP, cP, Z) \to 1 \mid Z = e(P, P)^{abc}] - \\ &- \Pr[A(P, aP, bP, cP, Z) \to 1 \mid Z - \text{випадкове}]|. \end{aligned} \quad (18)$$

Говорять, що DBDH-задача у (p, G, G, G_T, e) є (ε, t)-складною, якщо для будь-якого A щонайбільше за час t, $Adv^{DBDH}(A) \le \varepsilon$.

Як було відмічено раніше, якщо використовуються параметри (p, G, G, G_T, e) для спарювання Типу 1, тоді DDH-задача у G не є складною, для даних $(P, aP, bP, cP) \in G^4$, необхідно лише зробити перевірку $e(P, cP) \stackrel{?}{=} e(aP, bP)$, але це не означає, що CDH-задача у G теж не є складною. Фактично, на даний момент не існує методів розв'язання задачі CDH у G з використанням білінійного відображення e. Групи, для яких задача DDH є простою, але задача CDH – складною, називаються лакунарними групами Діффі – Геллмана, також існує відповідна лакунарна задача Діффі-Гелмана (GDH – gapDiffie-Hellman).

6) DLIN-задача (DecisionLinear)

Вхідні дані: Циклічна група $G = \langle P \rangle$ і набір (P, aP, bP, acP, bdP, Q).

Завдання: Визначити, чи є $(c+d)P=Q$, або, чи є Q випадковим рівномірним елементом G, який також є незалежним від інших елементів.

Задачі, що описані у даній статті, мають фіксоване число елементів G, відповідно, нестатичні задачі мають змінне число елементів G. Фактичне число елементів далі буде позначатися параметрами h і q.

7) BDHE-задача (BilinearDiffie-HellmanExponent)

Вхідні дані: Набір $(P,aP,a^2P,\ldots,a^{h-1}P,a^{h+1}P,\ldots,a^{2h}P,Q)$, де, $a\in Z_p$, $Q\in G$ – випадкові елементи і $G=\langle P\rangle$.

Завдання: Обчислити $e(P,Q)^{a^h}$.

8) DBDHE-задача (DecisionalBilinearDiffie-HellmanExponent)

Вхідні дані: Набір $(P,aP,a^2P,\ldots,a^{h-1}P,a^{h+1}P,\ldots,a^{2h}P,Q,Z)$, де, $a\in Z_p$, $Q\in G$ – випадкові елементи і $Z\in G_T$.

Завдання: Визначити, чи є $Z=e(P,Q)^{a^h}$, або, чи є Z випадковим рівномірним елементом G_T, який також є незалежним від інших елементів.

9) BDHI-задача (BilinearDiffie-Hellmaninversion)

Вхідні дані: Набір $(P_1,\ldots,P_h)$, де $P_i=a^iP$ для деякого випадкового $a\in Z_p$.

Завдання: Обчислити $e(P,P)^{1/a}$.

10) DBDHI-задача (DecisionalBilinearDiffie-Hellmaninversion)

Вхідні дані: Набір $(P_1,\ldots,P_h,Z)$, де $P_i=a^iP$ для деякого випадкового $a\in Z_p$ і $Z\in G_T$.

Завдання: Визначити, чи є Z еквівалентним $e(P,P)^{1/a}$, або, чи є Z випадковим рівномірним елементом G_T, який є незалежним від a.

11) wDBDHI-задача(WeakDecisionalBilinearDiffie-HellmanInversion)

Вхідні дані: Набір $(Q,P_1,\ldots,P_h,Z)$, де $P_i=a^iP$ для деякого випадкового $a\in Z_p$, $Q\in G$ – випадковий елемент, $Z\in G_T$.

Завдання: Визначити, чи є Z еквівалентним $e(P,P)^{1/a}$, або, чи є Z випадковим рівномірним елементом G_T, який є незалежним від a.

12) wDBDHI* – задача(друга версія попередньої задачі)

Вхідні дані: Набір $(Q,P_1,\ldots,P_h,Z)$, де $P_i=a^iP$ для деякого випадкового $a\in Z_p$, $Z\in G_T$, $Q\in G$ – випадковий елемент.

Завдання: Визначити, чи є Z еквівалентним $e(P,Q)^{h+1}$, або, чи є Z випадковим рівномірним елементом G_T, який є незалежним від a.

Обчислювальні версії останніх двох задач є еквівалентними над лінійно-часовим зведенням і за допомогою алгоритму рішення даних задач можна вивести алгоритм для обчислювальної версії важко зведеної BDHI-задачі [14].

13) q-ABDHE-задача (TruncatedDecisionalAugmentedBilinearDiffie-HellmanExponent)

Вхідні дані: Набір $(Q,\alpha^{q+2}P',P,\alpha P,\alpha^2P,\ldots,\alpha^qP,\alpha^{q+2}P,\ldots,\alpha^{2q}P)$.

Завдання: Обчислити $e(P,Q)^{\alpha^{q+1}}$.

Існує декілька версій цієї задачі:

а) Скорочена версія

Вхідні дані: Набір $(Q,\alpha^{q+2}Q,P,\alpha P,\alpha^2P,\ldots,\alpha^qP)$.

Завдання: Обчислити $e(P,Q)^{\alpha^{q+1}}$.

б) Вирішальна (decision) версія, запропонована Джентрі[15]

Вхідні дані: Набір $(Q,\alpha^{q+2}Q,P,\alpha P,\alpha^2 P,\ldots,\alpha^q P,Z)$.

Завдання: Визначити,чи є $Z=e(P,Q)^{\alpha^{q+1}}$, або, чи є Z випадковим рівномірним елементом G_T.

Далі, дамо визначення декільком задачам для спарювання Типу 2. Для визначення структури крупи на еліптичній кривій будемо використовувати спарювання Вейля. Для даного $N=\#E(F_q)$, якщо N не має квадратних коефіцієнтів, тоді структура групи $E(F_q)$ є ізоморфною до Z/NZ. Якщо r^2 ділиться на N, тоді повинна існувати точка порядку r^2, або дві незалежні точки порядку r. Для спарювання Вейля можуть існувати тільки дві незалежні точки, такі, що r ділиться $(q-1)$.

14) ECDDH-задача (TheEllipticCurveDecision Diffie-Hellman)

Вхідні дані: Набір точок P_1,P_2,P_3,P_4 у $E(K)[r]$, такі, що $P_2=[\lambda]P_1$, для деякого λ.

Завдання: Визначити, чи є $P_4=[\lambda]P_3$.

В загальному випадку ця задача на білінійних відображеннях може бути вирішена наступним чином. Якщо $e(P_1,P_3)\neq 1$, тоді достатньо перевірити рівняння $e(P_1,P_4)=e(P_2,P_3)$.

Жу і Нгуєн [16] запропонували алгоритм який працює, в тому числі, і на суперсингулярні кривій, що розв'язує вирішальну версію задачі Диффі – Гелмана на спарюваннях за поліноміальний час, але, вважається, що обчислювальна (computational) задача Діффі – Гелмана має рівень стійкості, що дорівнює задачі дискретного логарифмування. Тому, вважається, що ECDDH дійсно менш важка задача за ECCDH.

15) BDH-задача (Bilinear Diffie-Hellman)

Вхідні дані: Точки P, $Q=\psi([c]P)$, $P_1=[a]P$, $P_2=[b]P$ такі, що $e(P,Q)\neq 1$.

Завдання: Обчислити $e([ab]P,Q)$.

Вважається, що BDH-задача є більш важкою за ECDH- задачу (EllipticCurveDiffie-Hellman), а також за задачу Діффі – Гелмана над кінцевими полями. Так, якщо маємо DH-оракул над E, вхідними даними для обчислення $[ab]P$ є (P,P_1,P_2), необхідно обчислити $e([ab]P,Q)$. Якщо маємо DH-оракул над кінцевим полем, обчислюємо $h_1=e(P,Q)$, $h_2=e(P_1,Q)=h_1^a$, $h_3=e(P_2,Q)=h_1^b$, і на вхід оракула подаємо (h_1,h_2,h_3), тоді в результаті отримаємо $h_1^{ab}=e([ab]P,Q)$.

16) DBDH-задача (DecisionBilinear Diffie-Hellman)

Вхідні дані: Точки P, Q такі, що $e(P,Q)\neq 1$, $P_1=[a]P$, $P_2=[b]P$ і g.

Завдання: Визначити, чи є $g\overset{?}{=}e([ab]P,Q)$.

Дана DBDH-задача не є складнішою за вирішальну версію задачі Діффі – Гелмана над кінцевим полем.

У роботі [5] було показано, що можна ефективно розв'язати задачу Діффі – Гелмана над двома циклічними групами G і H простого порядку r, таких, що $G\times G\to H$, де $H\subset F_{q^k}^*$ – прообраз підгрупи, якщо існує ефективно обчислювальний гомоморфізм з H до G.

Існує ще декілька обчислювальних задач[17], що засновані на спарюваннях, в даній статті були наведені лише основні задачі, що частіше за інші використовуються в криптографічних системах на ідентифікаційних даних.

Висновок

Наведено класифікацію та аналіз основних обчислювальних задач на білінійних відображеннях, а також визначення та основні математичні відомості, що необхідні для розумінні даних задач.

Стійкість BDH-задач над еліптичною кривою $E(F_q)$, а також над кінцевим полем F_{q^k}, лежить від стійкості задачі Діффі – Гелмана. Найчастіше, для криптографічних перетворень, що засновані на білінійних відображеннях, використовується підгрупа над $E(F_q)$ достатньо великого простого порядку r. Мінімальним рівнем безпеки для наведених задач є $r > 2^{160}$ і $q^k > 2^{1024}$. Взагалі ефективність криптосистеми залежить від особливостей побудування усієї схеми, але, чим менше значення приймає q, тим швидше виконуються арифметичні операції над кривою $E(F_q)$. Тому необхідно, щоб параметр q приймав найменше значення, тоді основним параметром безпеки буде виступати k, яке повинне приймати якомога більші значення. Найчастіше обирається точки на $E(F_q)$, для яких $q \approx 2^{170}$, і еліптичні криві зі вкладеним ступенем $k = 6$, тобто $q^k \approx 2^{1024}$.

Для криптографічних систем на білінійних відображеннях основними питаннями є: як знайти підходящу еліптичну криву і як швидко обчислити $e(P,Q)$, тому важливим завданням є аналіз обчислювальних задач на решітках, а також об'єднаної математики алгебраїчних решіток і білінійних відображень, що, на наш погляд, дозволить прискорити криптографічні перетворення та покращити показники стійкості криптографічних систем на ідентифікаторах.

Список літератури:1. *Sanjit Chatterjee, Palash Sarkar.* Identity-Based Encryption / Springer Science + Business Media, LLC.– 2011. – P.29-30, 44-48, 125-135. 2. *J.H. Silverman.* The Arithmetic of Elliptic Curves / GTM.– Vol.106.– Springer-Verlag.– 1986. 3. *D. Lorenzini.* An Invitation to Arithmetic Geometry / AMS, Graduate Studies in Mathematics.– Vol.106.– 1993. 4. *Ишмухаметов, Ш.Т.* Методы факторизации натуральных чисел : учеб. пособие / Казанский университет. – 2011. – С.100-114. 5. *Ian F. Blake, GadielSeroussi, Nigel P. Smart.* Advances in Elliptic Curve Cryptography / London Mathematical Society Lecture Note Series. 317. – Cambridge University Press.– 2005.– P.183-211. 6. *S. Lichtenbaum.* Duality theorems for curves over p-adic fields / Inventiones Math.– Vol. 7.– 1969.– P.120–136. 7. *G. Frey, H.-G. Ruck.* A remark concerning m-divisibility and the discrete logarithm problem in the divisor class group of curves / Math.Comp. – Vol. 62.– 1994.– P.865–874. 8. *G. Frey, M. Muller, H.-G. Ruck.* The Tate pairing and the discrete logarithm applied to elliptic curve cryptosystems // IEEE Trans. Inf. Theory. – Vol. 45.– 1999.– P.1717–1719. 9. *F. Hess.* A note on the Tate pairing of curves over finite fields / Arch. Math.– Vol. 82.– 2004.– P.28–32. 10. *R. Balasubramanian, N. Koblitz.* The improbability that an elliptic curve has subexponential discrete log problem under the Menezes–Okamoto–Vanstone algorithm / J. Cryptology.– Vol. 11.– 1998.– P.141–145. 11. *T. Garefalakis.* The generalised Weil pairing and the discrete logarithm problem on elliptic curves / In S. Rajsbaum, edit. – Lecture Notes in Computer Science LATIN 2002 «Theoretical Informatics». – Vol.2286.– Springer-Verlag.– 2002.– P.118–130. 12. *Steven D. Galbraith, Kenneth G. Paterson, Nigel P. Smart.* Pairings for cryptographers / Discrete Applied Mathematics.– Vol. 156(16).– 2008.– P.3113–3121. 13. *Dan Boneh, Matthew K. Franklin.* Identity-based encryption from the Weil pairing / SIAM J. Comput.– Vol. 32(3).– 2003.– P.586–615. 14. *Dan Boneh, Xavier Boyen, Eu-Jin Goh.* Hierarchical identity based encryption with constant size ciphertext / In Ronald Cramer, edit. – Lecture Notes in Computer Science «Annual International Conference on the Theory and Applications of Cryptographic Techniques». – Vol.3494.– Springer.– 2005.– P.440–456. 15. *Craig Gentry.* Practical identity-based encryption without random oracles / In Serge Vaudenay, edit. – Lecture Notes in Computer Science «Annual International Conference on the Theory and Applications of Cryptographic Techniques». – Vol. 4004.– Springer.– 2006.– P.445–464. 16. *A. Joux, K. Nguyen.* Separating Decision Diffie–Hellman from Diffie–Hellman in cryptographic groups / J. Cryptology.– 2003.– P.239–248. 17. *A. Joux.* The Weil and Tate pairings as building blocks for public key cryptosystems / In C. Fieker, D.R. Kohel, edit. – ANTS-5 «Algorithmic Number Theory».– Vol.2369.– Springer.– 2002.– P.20–32.

Харьковский национальный университет радиоэлектроники

Поступила в редколлегию 15.09.2012

УДК 621. 391:519.2:519.7

Д.В. ІВАНЕНКО

МЕТОДИ ПРОТИДІЇ АТАКАМ СПЕЦІАЛЬНОГО ВИДУ НА СХЕМИ НАПРАВЛЕНОГО ШИФРУВАННЯ У КІЛЬЦЯХ ЗРІЗАНИХ ПОЛІНОМІВ

Вступ

Відсутність нових відкритих публікацій про вразливості та недоліки математичної моделі сучасних асиметричних криптосистем при існуванні відомих успішних реалізації на схеми направленого шифрування наводить на думку науковий світ, що зловмисники звернули увагу безпосередньо на програмну та апаратну реалізацію. Тобто завдання зловмисника зводиться до дослідження реалізованої криптосистеми з метою пошуку залежностей, вразливостей та недоліків. Гонка за швидкістю, великою продуктивністю за рахунок оптимізації, використання сучасних технологій веде до складностей при апаратній реалізації, людський фактор – до помилок у реалізації. Все це веде до імовірності реалізації атак [1] на реалізацію.

Такі атаки отримали назву атаки спеціального виду, існування таких атак формулює завдання аналізу існуючих та можливих методів протидії атакам спеціального виду.

В цій роботі проаналізовано методи протидії атакам, які базуються на аналізі енергоспоживання, на NTRU [2].

Існуючі методи протидії атакам на реалізації

З появою атак спеціального виду виникла задача створити методи та засоби протидії цим атакам. Наведемо класифікацію та дамо визначення наступним методам протидії:

1) Використання фіксованих схем операції для кожної моделі операції:

а) Метод Монтгомері [3 – 6]. Цей метод дозволяє прискорити операцію множення та зведення у квадрат;

б) Доповнення до операцій [7 – 10]. До основної групи операції додаються змішані додавання, фіксовані додавання, єдині додавання, подвоєння;

в) Атомарність [11]. Це властивість операцій під якою розуміється, що операція виконується як єдине ціле або не виконується взагалі. Тобто, внаслідок порушення процедури виконання операції процедура виконання відкатується до початку операції, і всі операнди операції генеруються спочатку.

2) Рандомізація даних [12];

3) Засліплення даних [13 – 15]. Засліплення служить для зміни вхідних даних в деякий непередбачуваний стан. Залежно від характеристик функції засліплення вона може виключити деякі або всі витоки корисної інформації.

Вибір метода протидії ускладнюється ще тим, що атака спеціального виду є атакою часного випадку. Тобто успішною реалізацією на одному пристрою може бути неефективною для іншого пристрою.

Також при впроваджені метода протидії потрібно враховувати особливість системи, в яку вбудовується. Для того щоб реалізація метода протидії не впливала на ефективність та на властивості за якими, було обрано криптоперетворення.

Криптосистема з відкритим ключем NTRU

Для аналізу методів протидії атакам спеціального виду обрали схему направленого шифрування у кільцях зрізаних поліномів, NTRU. Наш вибір був оснований на аналізі сучасних криптосистем [16].

Алгоритм включає в собі три відкритих параметри (N,q,p), де N,q,p – цілі, та q,p – взаємно прості при тому, що p значно менше від q. Процедура генерації ключа починається з вибору випадкового $F \in R$ та вирахування $f:=1+pF$. У випадку, якщо f не має зворотного елемента за $mod\ q$, то потрібно обрати інший F та повторити процедуру. Далі необхідно виб-

рати другий поліном $g \in R$, який буде зворотній за *mod q*. Таким чином, f – це секретний ключ, а h – відкритий ключ, який потрібно обчислити з наступного виразу: $h:=pf^{i} * g \bmod q$. Повідомлення позначається через m. При шифруванні відправник випадково вибирає поліном $r \in R$, потім обчислює значення шифр-тексту $e:=r*h+m \bmod q$. Для розшифрування e отримувач обчислює $a:=f*e \bmod A\ q$, де *modA q* означає, що коефіцієнти зсуваються на інтервал $[A, A+q-1]$ після приведення за *mod q*. Потім отримувач відновлює текст за формулою – $m:=a \bmod p$.

Домінуючою операцією в шифруванні та розшифруванні NTRU вважається обчислення $r*h \bmod q$ та $f*e \bmod q$. Коефіцієнти поліномів h та e розподілені майже випадково, один вибирає r та F так, що результат операції згортання $r*h$ та $F*e$ може бути мати більш низьку обчислювальну складність. В алгоритмі використовуються бінарні або тернарні коефіцієнти, $r_i, F_i \in \{0,1\}$ або $\{-1,0,1\}$, і фіксується число ненульових коефіцієнтів у r та F [17, 18]. Потім обчислення операції згортання може бути обчислено за dN-операції, де N – загальне число коефіцієнтів поліномі та d – число не нульових коефіцієнтів.

Алгоритм 1 представляє собою оптимізований алгоритм обчислення $t=a*c \bmod q$, де зниження за модулем *mod N* пересуває в індексі обчислення по масиву t за рахунок додаткової пам'яті $(t_N, .., t_{2N-1})$. В цьому алгоритмі $a \in R$ є бінарним поліномом з d ненульовими коефіцієнтами та $c \in R$ – загальним поліномом. Бінарний поліном a представляє собою масив b, у якому позначено ненульові положення d. Наприклад, $a(X)=1+x3+x6= =[1,0,0,1,0,0,1,0]$ для $N=8$ буде мати наступний вигляд $b=[0,3,6]$.

Наприклад: розглянемо ситуацію, де зловмисник намагається відновити секретний ключ ($a(x)$), при тому що він контролює вхідні дані (поліном c(x)) та знає результат операції згортки ($t(x)$). Припущення відносно знання секретного ключа зловмисник робить за допомогою аналізу спектру енергоспоживання при виконанні операції згортки на пристрої.

Алгоритм 1 – Обчислення операції згортання[19].

Вхідне: масив b, який представляє собою ненульові значення бінарного поліному $a(X)$, що є секретним ключем); поліном $c(X)$.

Вихідне: $t(x)$ тимчасовий буфер, у якому зберігається результат.

1. for $0 \le j < 2N$ do
2. $t_j \leftarrow 0$ //з t_N до t_{2N-1} : тимчасовий буфер
3. end for
4. for $0 \le j < d$ do
5. for $0 \le k < N$ do
6. $t_{k+b[j]} \leftarrow t_{k+b[j]} + c_k$
7. end for
8. end for
9. for $0 \le j < N$ do
10. $t_j \leftarrow (t_j + t_{j+N}) \bmod q$
11. end for

Запропоновані методи протидії атакам спеціального виду на енергоспоживання

Аналіз алгоритму NTRU показує, що головною математичною операцією під час зашифрування та розшифрування у алгоритмі NTRU є добуток двох поліномів $a(X)$ та $b(X)$ та операція згортання $t(X)=c(X)+a(X)$, де $a(X)$ представлений у вигляді множини b. Тому було би ефективно запропонувати методи протидії, пов'язані саме з цими операціями:

1. Рандомізація тимчасових даних, які зберігаються у t;
2. Засліплення відкритих даних c;
3. Рандомізація секретних даних b.

Мета цієї статі полягає не тільки у аналізі запропонованих методів протидії атакам спеціального виду, а ще й у знаходженні таких методів, реалізація яких не впливала би на ефективність самого алгоритму, тобто щоб алгоритм не втрачав своєї початкової швидкості, стійкості, ефективності.

Загальним принципом впровадження методів є рандомізація та складність вияву залежності операції над значенням секретних даних, важливим аспектом реалізації кожного методу протидії є необхідність підтримувати початкові властивості алгоритму: швидкість, стійкість, ефективність.

Використовуючи класифікацію атак спеціального виду на енергоспоживання [1], проаналізуємо запропоновані методи протидії для криптосистеми NTRU.

Рандомізація тимчасових даних, які зберігаються у *t*

Першим методом протидії є ініціалізація *t* з випадковими цілими числами, а не нулями. Ця протидія направлена як на SPA, так і на CPA(включаючи DPA). Ініціалізація *t* з невідомими значеннями ускладнює для зловмисника реалізацію SPA та робить складнішим пошук кореляції між енергоспоживанням та основним додаванням, тому що один з двох операндів має випадкове значення. Тобто імовірність знаходження максимумів, при аналізі спектру енергоспоживання зловмисником, йде до нуля, що і потрібно було зробити.

На рис. 1 показано максимальне значення коефіцієнта кореляції після використання першого заходу протидії, з цього можна зробити висновок, що пошук максимального піку для визначення значення *b[l]-b[l-1]* є досить складним.

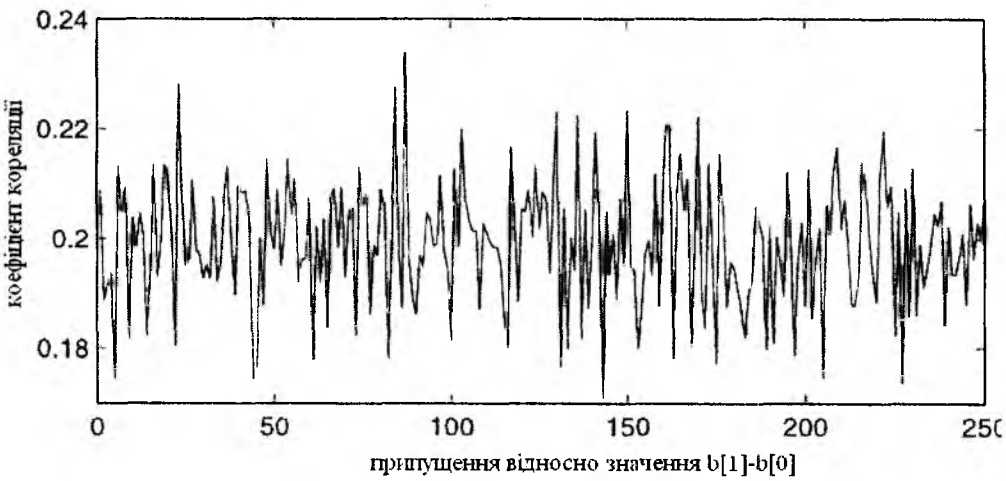


Рис. 1. Максимальна кореляція для кожного можливого припущення після впровадження протидії №1

Для того щоб реалізація алгоритму не втратила необхідну ефективність, необхідно щоб $t_{\max} \le d \times \max_{0 \le k < N} c_k$, де t_{max} – максимум серед t при $0 \le j \le 2N$.

Таким чином, впровадження цього заходу тільки обмежить стійкість алгоритму до CPA, так як обмежується вибір r_j. Наприклад, шукаючи вихідне значення *b[1]-b[0]=3* (див.рис.1), зловмисник має $HD(c_3, c_3+c_0)$, та це значення зміниться на $HD(r_4+c_3, r_4+c_3+c_0)$, у випадку реалізації першої протидії. Тоді буде безглуздо використовувати r_4, який набагато більший ніж c_3+c_0, тому що з високою імовірністю положення старшого біту в r_4 майже не сприяє знаходженню відстані Хеммінга для цього c_3+c_0. Тому розумно припустити, що r_4 буде лежати в інтервалі $[0, c_3+c_0]$. Просте узагальнення цього спостереження для інших значень *b[l]-b[l-1]* призводить до рішення, що розумно припустити, що $r_j \le t_{\max}$. Якщо $\max_{0 \le k < N} c_k$ у звичайному шифр-тексті, то $r_j < dq$.

Після цього зловмисник може зробити припущення щодо деякого фіксованого значення всіх r_j при $r < dq$ та спробувати розпочати схожу атаку до оригінального CPA. Такий вид

атаки може визначити як потенційний, тому що у випадку, коли зловмисник накопив достатнє число слідів енергоспоживання, він буде мати багато випадків, де $r_{b[l]}=r$ і випадок, коли всі інші стани будуть відігравати роль шуму. Однак виникає проблема, що кожне r_j буде впливати тільки на одне t_j протягом виконання алгоритму. Таким чином, якщо наведена вище атака успішна для розкриття *b[1]-b[0],* то вона також може бути використана для знаходження всіх значень *b[l]-b[l-1],* без істотного збільшення числа потрібних слідів енергоспоживання.

Засліплення відкритих даних *c*

Розглядаючи цей метод можна запропонувати два варіанта цього методу: рандомізація цілих чисел та рандомізація поліному. Перший реалізує рандомізацію цілого *r* та неодноразово використовується для засліплення всіх $c_к$. Цей захід протидії маскує кореляцію між поліномом *c(X)* та енергоспоживанням рандомізованого *c(X).* Насправді, ця процедура має наступний вигляд *(c(X)+R(x))*a(X)-R(X)*a(X),* де *R(X)=[r,r...,r]*. Тому що *R(X)*a(X)=[dr mod q,dr mod q,...,dr mod q]*, можемо ліквідувати *R(X)*a(X)* шляхом віднімання dr від кожного t_j

Цей метод можна назвати досить прийнятним за рахунок незначних накладних витрат. Однак потрібно бути обережними при використані засліплення у ситуації, коли енергоспоживання обчислюється за допомогою коду Хеммінга, тому що інший вид атаки CPA теоретично буде можливим в цьому випадку. Наприклад, вважатимемо, що енергоспоживання змінюється від найменш значущого біту кожного t_j. Спочатку пояснимо атаку CPA проти алгоритму без використання протидій. Як зазначалося вище, завдання зловмисника полягає у знаходженні вихідного відносного зсуву індексу *t*, яке показує *b[1]- b[0], b[2]-b[1]* і так далі. Візьмемо поліном $c^1,...,c^S$ та його сліди енергоспоживання $P^1,...,P^S$ зробимо припущення, що *b[1]- b[0]=w* при $w \geq 1$ та перевіримо на наявність помітних кореляцій між $LSB(c_w^l + c_0^l)$ та P^l, де *LSB(x)* це значення найменш значущого біту *x*. Цікаво, що ця атака буде реалізована навіть якщо *r* додане до кожного $c_к$, $LSB((c_w^l + r) + (c_0^l + r)) = LSB(c_w^l + c_0^l)$. На рис. 2 показано, що пік з'являється серед коефіцієнтів кореляції, хоча це важко зробити на рис. 3 тому, що енергоспоживання нашого процесору впливає більш швидше на відстань Хеммінга, ніж на код Хеммінга.

Решта завдання атаки, відновлення *b[2]-b[1], b[3]-b[2]* і так далі ускладнюється в результаті втрати залежності *r* на *LSB.* Ця проблема може буде вирішена припущенням двох відносних зсувів в той же час. Вважатимемо, що зловмисник відновив *b[1]- b[0]=v* використовуючи вище зазначений метод. Після чого зловмисник робить припущення *b[3]- b[1]=*w_1 та *b[3]-b[1]=*w_2 та знаходить кореляцію використовуючи $LSB(c_{w_1+v} + c_{w1} + c_{w2} + c_0)$. Наприклад, *v=3,* w_1*=3,* w_2*=2,* (див. рис. 2 [3]). Цей метод успішно ліквідовує ефект засліплення для *r* тому, що значення, з яким зловмисник має справу, складається з парного числа *r*. Хоча простір пошуку (w_1, w_2) буде більшим, ніж просте *w,* він все ще знаходиться в практичному діапазоні. Тому, розглядаючи перший метод з простим цілим *r* , можна вважати потенційним рішенням для вирішення задачі рандомізації відкритих даних.

Вирішити цю задачу може і наступний метод, він заснований на використані загального поліному *R(X),* який безпосередньо і буде засліплювати *c(X).* Вирахуємо *(c(X)+R(X))*a(X)-R(X)*a(X)=c(X)*a(X)* використовуючи випадковий поліном *R(X).* Для зменшення навантаження обчислення *R(X)*a(X)* взяли за ідею [5, 6]. Тобто спочатку обирається випадкове *R(X),* обчислюється *S(X)=R(X)*a(X),* та зберігається *R(X)* та *S(X).* Коли згортання *(c(X)+R(X))*a(X)-S(X)* буде зроблено, *R(X)* та *S(X)* оновлюється шляхом обчислення $R(X) \leftarrow kR(x)$ та $S(X) \leftarrow kS(x)$ для випадкового *к.* Тому можна використовувати другий метод як ефективну протидію проти CPA. На рис. 3 показано максимальне значення коефіцієнта кореляції після застосування цієї протидії. Але все одно знаємо, що засліплення *c* не перешкодить реалізації SPA атаки.

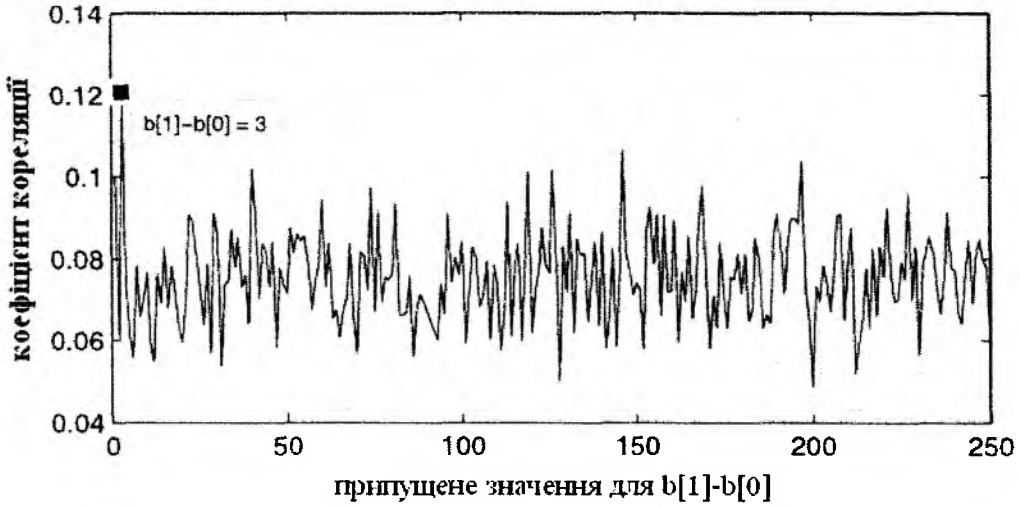


Рис. 2. Максимальна кореляція для кожного можливого припущення при моделі коду Хеммінга з *c*, який засліплений випадковим цілим числом

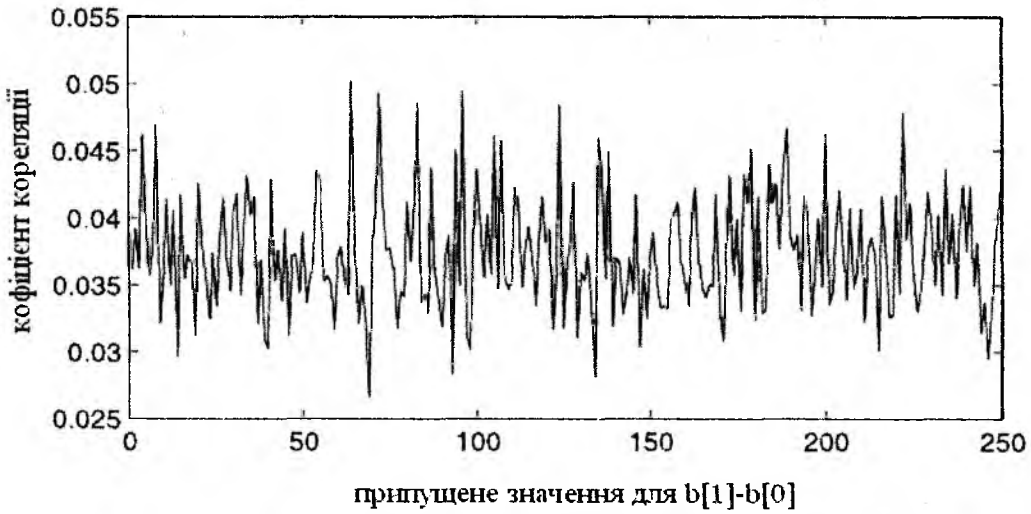


Рис. 3. Максимальна кореляція для кожного можливого припущення з *c*, який засліплений випадковим поліномом

Висновки відносно атаки другого роду

Проаналізувавши запропонований захід протидії, а саме рандомізацію вектора r_k, зловмисник може зробити висновок про ймовірність реалізації атаки CPA. Тобто, є ймовірність того, що рандомізація вектору r_k можна відокремити від сигналу енергоспоживання для отримання корисної інформації.

Спочатку розглянемо сутність атаки CPA. Завдання зловмисника полягає у відновленні *b[1]-b[0]* для отримання множини *b*. Для цього зловмисник описує вираз енергоспоживання використовуючи модель відстані Хеммінга(1), робить припущення *P* відносно значення енергоспоживання, тобто значення регістру, що змінюється підчас виконання операції згортання поліномів:

$$P(y \to z) \approx mHD(y,z) + n = mHW(y \oplus z) + n \qquad (1)$$

де m – скалярний коефіцієнт посилення між відстанню Хеммінга та енергоспоживанням і n – незалежна зміна шуму.

Повернемося до рис. 1 [3], де розглянуто приклад, за яким зловмисник робить припущення *w* відносно значення *b[1]-b[0]*. Його припущення будується на відстежуванні відмін-

ностей кореляції між $P(r_k \to r_k + c_w)$ та $HD(r_k, r_k + c_w)$ при $j=0$ та кореляції між $P(r_k + c_w \to r_k + c_w + c_0)$ та $HD(r_k + c_w, r_k + c_w + c_0)$ при $j=1$. Усі припущення зловмисника можна представити у вигляді

$$HW(x \oplus (x + y)) \tag{2}$$

та привести до

$$HW(x \oplus (x + y)) = HW(y) + \epsilon, \tag{3}$$

де ϵ – імовірнісна залежність зсувів при операції згортання поліномів з використанням рандомізації вектора r_k.

Таким чином, у випадку, коли зловмисник отримав достатньо статистичних даних про операції згортання на деякій множині поліномів, він, проаналізувавши енергоспоживання отриманих операцій згортання та використовуючи вирази (1) та (2), може припустити, що можна знайти залежність зсувів при використанні рандомізованого вектору r_k. Це припущення сформоване на підставі складності вибору r_k та переповненні або переносі регістра зсуву при додаванні r_k до добутку поліномів(див. (3)). Потім зловмисник відфільтровує шум, маючи на увазі під ним корисну інформацію про рандомізований вектор r_k, та отримує сигнал даних енергоспоживання операції згортання без реалізації заходів протидії атакам на енергоспоживання, то проводить атаку спеціального виду [3], тим самим відновлюючи секретний ключ.

Рандомізація *b*

Під час спостережень за поведінкою алгоритму 1 виявився цікавий факт, що зміна порядку елементів множини b не впливає на результат операції згортання. Цей факт відкриває ефективність використання рандомізації b, як міри протидії атаці CPA. Ця протидія може ефективно запобігти CPA, тому що CPA використовує позиції відносних зсувів, які додаються до c_0 між двома послідовними операціями в головному циклі, але мета порівняння та значення цих зсувів змінюється цією протидією при кожному виконані. Наприклад, при ітерації $j=0$ та $j=1$, відносний зсув більше не дорівнює b[1] –b[0], крім того він більш немає фіксованого значення. На рис. 4 показана максимальне значення коефіцієнтів кореляції після застосування рандомізації b.

Проте необхідно відмітити, що рандомізація b може мати потенційну слабкість до SPA атаки, якщо зловмисник зробить декілька успішних припущень.

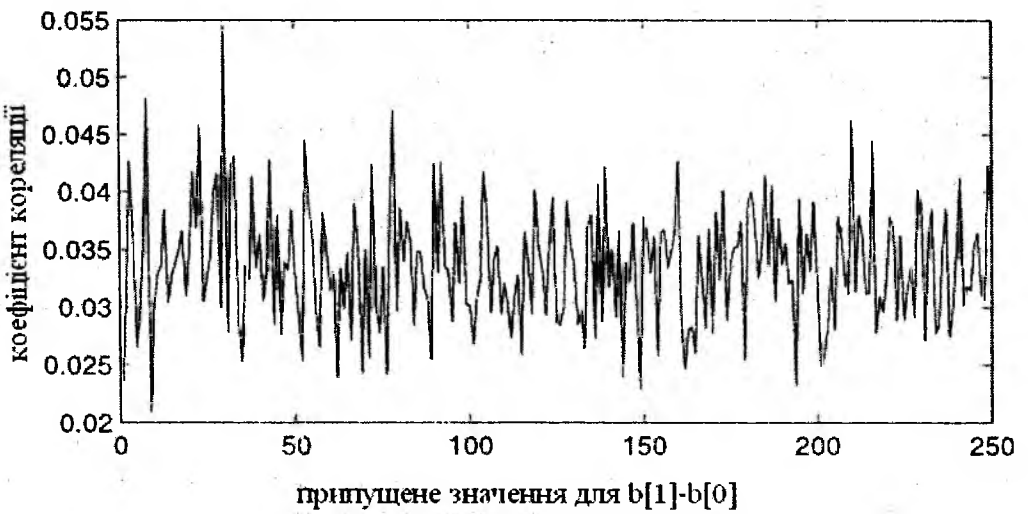


Рис. 4. Максимальна кореляція для кожного можливого припущення з рандомізованим b

Нехай множина *b'[j]* є елементом в *j* позиції *b* після перестановки. Тоді легко помітити, що пара *(b'[1],b'[0])* може мати *d(d-1)* варіантів, тобто 2^{256} варіантів при *d=48*. Якщо зловмисник зібрав достатнє число слідів енергоспоживання, то всі *d(d-1)* варіанти також повинні бути включені до зібраних даних.

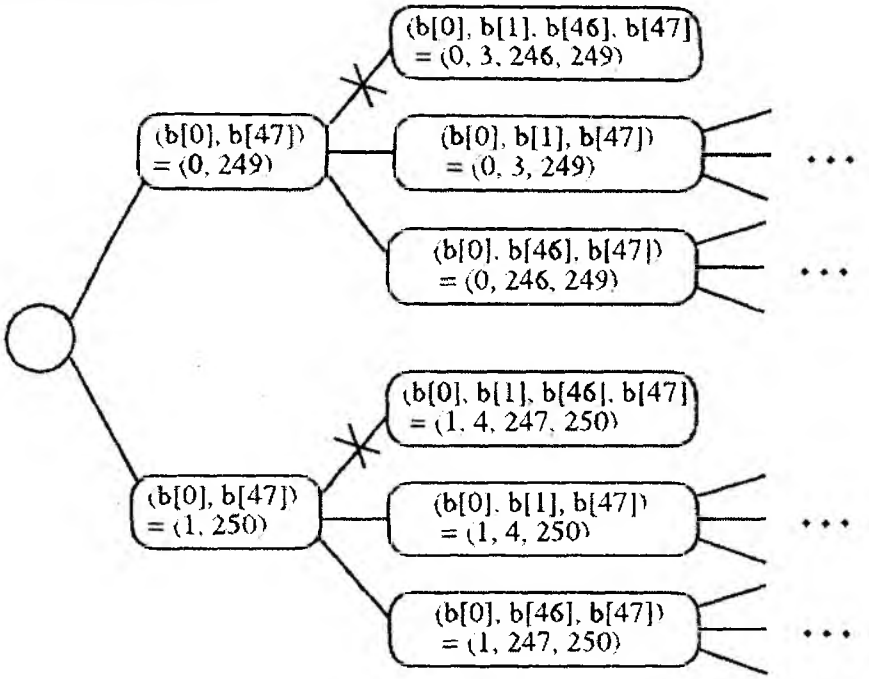


Рис.5. Приклад дерева пошуку при B=[…,239,242,244,245,246,249]

Потім зловмисник представляє атаку SPA незалежно для кожного з слідів енергоспоживання, відновлюючи упорядковану множину *B* усіх можливих значень при відносному зсуві *b'[1] –b'[0]*. Хоча ця множина *b* неявно показує оригінальну множину *b*, але може показати деяку частину інформації про *b*, якщо *b* буде вибрано необережно. Наприклад, будемо вважити, що при *N=251* та *d=48* останні елементи множини *B* буде мати наступний вигляд [239,242,244,245,246,249]. Потім останній елемент 249 інформує зловмисника, що *(b[0],b[47])=(0,249)* або *(1,250)*, тому що максимальний зсув має відбутися між *b[1]* та *b[d-1]*. Це значить, що зловмисник відновить два вихідних елемента з 48 з імовірністю 0,5, та ця процедура може бути продовжена для визначення інших елементів, сконструювавши дерево пошуку (див. рис.5). Для простоти пояснимо тільки першу гілку, тому що інша буде симетричною. Другий останній елемент у *B*, 246 породжує три дочірні гілки у дереві пошуку. Тобто *(b[0],b[1],b[46],b[47])=(0,3,246,249);* *(b[0],b[1],b[47])=(0,3,249)* та *b[46]≠246* *(b[0],b[1],b[46])=(0,3,246)* та *b[1]≠3*. Але перша гілка може бути скорочена, тому що перша гілка не входить до множини *B* у точці, в якій зловмисник не може знайти відносний зсув *b[46]-b[1]=243* у *B*. Виходячи з цього на наступному рівні, зловмисник отримує чотири секретних елемента з 12 варіантів, що істотно знижує пошук у просторі *b*. Хоча такий вид атаки буде корисний тільки при невеликій глибині дерева, тому що утворюється велика кількість гілок після певного рівня і обрізка практично неможлива, тому потрібно бути обережним при використання такої протидії проти атак SPA.

Аналіз продуктивності

Аналізуючи методи протидії, доцільно було б порівняти алгоритм [1] та алгоритми з різними методами протидії. На нашу думку, ці реалізації можна порівняти, аналізуючи використану пам'ять та швидкість операції обчислення, тобто, доказуючи ефективність протидії до розглянутих атак, потрібно також враховувати вплив реалізації методів на властивості алгоритму.

Розглянувши табл. 1, можна проаналізувати продуктивність різних реалізації щодо виконання операції згортання $t(X)=c(X)*a(X)$, яка є вразливою до атак спеціального виду на енергоспоживання. Як видно з таблиці, що засліплення *c* з випадковим поліном вимагає занадто багато пам'яті в порівнянні з іншими заходами протидії та більше використовує обчислювальний ресурс ніж інші заходи. Засліплення *c* з випадковим числом показує найвищу швидкість серед всіх заходів, але цей метод вразливий до атак типу CPA. Далі з таблиці видно, що порівнюючи використання пам'яті та швидкості обчислення можна виділити реалізацію одночасно двох заходів: рандомізація *t* та рандомізація *b*, реалізація яких зведе нанівець можливість CPA та SPA.

Таблиця 1

Алгоритм згортання	ROM	RAM	Затрачений час
Алгоритм 1 (без застосування протидії)	3796	1063	67.383
Рандомізація *t*	3980	1063	71.191
Засліплення *c* з цілим *r*	4022	1064	69.043
Засліплення *c* з поліномом *R(X)*	4194	2067	101.953
Рандомізація *b*	4050	1063	70.117
Рандомізація *t* та рандомізація *b*	4106	1064	73.828

Висновок

Таким чином, проаналізувавши методи протидії, виявили, що кожний метод має як переваги, так і недоліки, і назвати універсальний метод протидії, який би протидіяв однаково ефективно атакам спеціального виду, виявилося складно. Насамперед, методи ускладнюють завдання зловмисника з відновлення секретного ключа, але при реалізації методів протидії потрібно значну увагу приділять впливу цих методів на сам алгоритм.

Рандомізація *t* ускладнює SPA та робить складнішим пошук кореляції, але обмежується вибором r_j, що може призвести до потенційних атак. Засліплення *c* може реалізовуватися двома варіантами, нести незначні накладні витрати, але вразливе до атак, які використовують для аналізу даних код Хеммінга та при реалізації засліплення *c* за допомогою поліному потрібно більше RAM(вдвічі більше ніж у інших випадках). Рандомізація *b* ефективна проти CPA, але неефективна проти SPA, при атаці SPA – контролюючи вхідні дані та використовуючи дерево пошуку, досить легко знайти залежності та досягнути мети.

Список літератури: 1. *Іваненко, Д.В.* Класифікація атак спеціального виду на енергоспоживання / Д.В. Іваненко // Системи обробки інформації. – Х. : ХУПС, 2012. – Вип.5(103). 2. *Hoffstein, J.* NTRU: A new high speed public key cryptosystem /J. Hoffstein, J. Pipher, J. Silverman Preprint // presented at the rump session of Crypto 96. 3. *Okeya, K.* Power analysis breaks elliptic curve cryptosystems even secure against the timing attack / K.Okeya, K. Sakurai // Indocrypt 2000, LNCS, vol.1977, pp.178–190, Springer, 2000. 4. *Brier, E.* WeierstraЯ elliptic curves and side-channel attacks / E. Brier, M. Joye // Public Key Cryptography – PKC 2002, LNCS, vol.2274, pp.335–345, Springer, 2002. 5. *Fischer, W.* Parallel scalar multiplication on general elliptic curves over Fp hedged against non-differential side-channel attacks / W. Fischer, C. Giraud, E. Knudsen, J. Seifert // 2002. International Association for Cryptologic Research (IACR) Cryptology ePrint Archive 2002/007, http://eprint.iacr.org/2002/007. 6. *Izu, T.* A fast parallel elliptic curve multiplication resistant against side channel attacks / T. Izu, T.Takagi // Public Key Cryptography – PKC 2002, LNCS, vol.2274, pp.280–296, Springer, 2002. 7. *Okeya, K.* Elliptic curves with the montgomery-form and their cryptographic applications / K. Okeya, H. Kurumatani, K. Sakurai // Public Key Cryptography 0 – PKC 2000, LNCS, vol.1751, pp.238–257, Springer, 2000. 8. *Liardet, P.* Preventing SPA/DPA in ECC systems using the Jacobi form / P. Liardet, N. Smart // Cryptographic Hardware and Embedded Systems – CHES 2001, LNCS, vol.2162, pp.391–401, Springer, 2001. 9. *Joye, M.* Hessian elliptic curves and side-channel attacks / M. Joye. J. Quisquater // Cryptographic Hardware and Embedded Systems – CHES 2001, LNCS, vol.2162, pp.402–410, Springer, 2001. 10. *Bernstein, D.* Faster addition and doubling on elliptic curves / D.Bernstein, T. Lange // Advances in Cryptology – Asiacrypt 2007, LNCS, vol.4833, pp.402–410, Springer, 2007. 11. Chevallier-Mames, B. Low-cost solutions for preventing simple side-channel analysis: Side-channel atomicity / B. Chevallier-Mames, M. Ciet, M. Joye // IEEE Trans. Comput., vol.53, no.6, pp.760–768, 2004. 12. *Coron, J.S.* Resistance against differential power analysis for elliptic curve cryptosystems / J.S.Coron // Cryptographic Hardware and Embedded Systems — CHES'99, LNCS,

vol.1717, pp.292–302, Springer, 1999. 13. Oswald, E. Randomized addition-subtraction chains as a countermeasure against power attacks / E. Oswald, M. Aigner // Cryptographic Hardware and Embedded Systems — CHES 2001, LNCS, vol.2162, pp.39–50, Springer, 2001. 14. *Moeller, B.* Securing elliptic curve point multiplication against side-channel attacks / B. Moeller // Information Security – ISC 2001, LNCS, vol.2200, pp.324–334, Springer, 2001. 15. *Hasan, M.* Power analysis attacks and algorithmic approaches to their countermeasures for Koblitz curve cryptosystems / M. Hasan // IEEE Trans. Comput., vol.50, no.10, pp.1071–1083, 2001. 16. *Іваненко, Д.В.* Порівняльний аналіз сучасних асиметричних криптосистем / Д.В. Іваненко, О.В. Сєвєрінов // Системи управління, навігації та зв'язку. – К. : ЦНДІ НіУ, 2012. – Вип.2(22). 17. *Lee, M.K.* Sliding window mothod for NTRU / M.K. Lee, J.W. Kim, J.E. Song, K. Park // Applied Cryptography and Network Security – ACNS 2007, LNCS, vol.4521, pp.432–442, Springer, 2007. 18. *Gama, N.* Symplectic lattice reduction and NTRU / N. Gama, N. Howgrave-Graham, P. Nguyen // Eurocrypt 2006, LNCS, vol.4004, pp.233–253, Springer, 2006. 19. Hoffstein, J. Optimizations for NTRU / J. Hoffstein, J. Silverman // Proc.Public-Key Cryptography and Computational Number Theory,2000.

Харківський національний
університет радіоелектроніки

Надійшла до редколегії 17.09.2012

УДК 681.3.06

Ю.И. ГОРБЕНКО, канд. техн. наук, Р.И. МОРДВИНОВ

СУЩНОСТЬ И АНАЛИЗ КРИПТОГРАФИЧЕСКИХ ТРЕБОВАНИЙ СТАНДАРТА NIST SP 800-90B

К источникам энтропии и методам генерации ключей в криптографии уделяется особое внимание, требования к ним изложены в ряде стандартов [1 – 4]. В то же время указанные стандарты относительно криптографии носят приложенческий характер. Поэтому важной задачей является, с одной стороны, формализация их требований, а с другой – рассмотрение ряда новых требований как к источникам энтропии, так и непосредственно к самим средствам генерации ключей.

Решение указанных задач представлено в рабочей версии стандарте NIST DRAFT Special Publication 800-90B [5].

Цель настоящей статьи – рассмотрение и анализ основных требований к источникам энтропии и непосредственно генераторам случайных последовательностей, стратегии их применения, возможности нестандартного поведения источников энтропии и требования валидации, а также общей стратегии и частных вопросов тестирования источников энтропии и генераторов ключей. Дополнительной целью является анализ и обоснование возможностей гармонизации этого стандарта в Украине.

Требования к источнику энтропии

Ниже, на основе [5], приведены основные требования, выдвигаемые к источнику энтропии, как в целом, так и к отдельным его компонентам. Данные требования должны помочь разработчикам в проектировании источников энтропии, способных удовлетворить требованиям, выдвигаемым в документации относительно их валидации.

Основные функциональные требования к источнику энтропии следующие [5]:

1. Разработчик должен документировать конструкцию источника энтропии в целом, включая взаимодействие компонентов. Данная документация должна обосновывать, почему источник энтропии способен вырабатывать биты с заданной энтропией.

2. Источник энтропии должен иметь четкую и понятную границу безопасности, которая должна быть такой же, либо находится в пределах криптографических границ приведенных в FIPS 140. Данная граница безопасности должна быть задокументирована и включать в себя:

- описание содержания границы безопасности, причем граница безопасности может выходить за пределы энтропии самого источника, т.е. источник энтропии может содержаться в больших границах неопределенности, чем детерминированный генератор случайных бит (ДГСБ), а граница безопасности может быть логической, т.е. не физической;
- описание того, как граница безопасности гарантирует, что нарушитель не сможет уменьшить энтропию, если он находится за пределами границы, либо путем наблюдения либо манипуляции;
- любые предложения относительно поддержки функций, от которых зависит граница безопасности.

3. Разработчик должен задокументировать диапазон рабочих условий, при которых источник энтропии может генерировать приемлемые случайные данные, причем должны быть четко описаны меры, которые были приняты при проектировании системы, с целью работы источника энтропии в нормальных условиях.

4. Предусмотрена возможность проверки источника энтропии на соответствие FIPS 140, включая соответствующие интерфейсы для получения данных.

5. Предоставлена документация, которая описывает поведение источника шума и объясняет, почему считается, что значение энтропии не колеблется при нормальных операциях.

6. После выявления в процессе тестирования неисправности источник энтропии должен прекратить вывод данных и уведомить приложение о состоянии ошибки.

7. Для улучшения устойчивости по отношению к деградации или некорректному поведению, источник энтропии может содержать несколько источником шума. Если данная функция реализована, то требования применяются к каждому из источников энтропии.

К полному(законченному) источнику энтропии выдвигаются еще и такие требования:

1. Выходная строка бит, полученная из источника энтропии, должна обеспечивать, по крайней мере, $(1-\varepsilon)n$ бит энтропии, где n – длина каждой выходной строки и $0 \le \varepsilon \le 2^{-64}$;

2. Выход источника энтропии должен формироваться с использованием криптографической функции, по крайней мере, двойной размер блока базового криптографического примитива должен быть использован в качестве входных данных в функцию состояния

Анализ требований к источнику шума

К источнику шума выдвигаются следующие функциональные требования [5]:

1. Источник шума должен формировать случайные выходные данные, т.е. на выходе не должно быть последовательностей с известными правилами;

2. Разработчик должен задокументировать функционирование источника шума. Документация должна включать в себя описание того, как источник шума работает, и обоснование, почему обеспечивается приемлемая выходная энтропия, ссылаясь на соответствующие исследования и литературу.

3. Для обеспечения правильной работоспособности источник шума должен иметь средства тестирования правильной работоспособности. В частности, должна быть обеспечена возможность сбора от источника шума данных для тестирования его работоспособности и определения значения энтропии, а также соответствие теста работоспособности. Также выходные данные, полученные от источника шума, должны обеспечивать свойства не оборотности, однородности и не влиять на последующие выходные;

4. Должна обеспечиваться возможность обнаружения отказа или серьезной деградации источника шума. Методы, используемые в таких условиях, должны быть задокументированы.

5. В документации на источник шума должны быть описаны условия, если таковые имеются, при которых известны неисправности, включая описание различных сред, в которых источник шума может работать неправильно. В этом случае непрерывное тестирование или другие механизмы источника энтропии должны обнаруживать неисправности;

6. Источник шума должен быть защищен, насколько это возможно, от ознакомления или влияния нарушителя. Методы, используемые для этого, должны быть задокументированы, в том числе описано использование границ безопасности для защиты источника шума от наблюдения или влияния со стороны нарушителя.

Особенности требований к валидации

Под валидацией понимается подтверждение на основе представления объективных свидетельств того, что требования, предназначенные для конкретного использования или применения, точно и в полном объёме предопределены, а цель достигнута. Проверка источника энтропии, т.е. его валидация, необходима для того, чтобы получить уверенность в том, что все соответствующие требования рекомендации выполнены. Проверка состоит из тестирования аккредитованной лаборатории источника шума в соответствии с требованиями SP 800-90B.

Проверка энтропии источника является сложной проблемной задачей. Это связано с тем, что ни одна другая часть генератора случайных бит (ГСБ) не является настолько зависимой от технологических различий и различий окружающей среды. В то же время, правильная работа источника энтропии является важной для безопасности ГСБ, и как следствие безопасности криптографического приложения.

В этом разделе представлены требования высокого уровня как для разработчиков, так и тестировщиков, которые должны быть выполнены для проверки методов отображения больших выборок в более мелкие битовые строки в ситуациях, когда сложно собрать достаточно данных для проверки с учетом размера выборки. Также требования, представленные ниже, предназначены для обеспечения проверки источника энтропии и обоснования, почему источник энтропии при их выполнении можно использовать.

Средства создания энтропия источника должны состоять, как минимум из трех компонентов: источника шума, средства тестирования работоспособности и дополнительных компонентов анализа состояния. Источник энтропии должен обеспечивать энтропию, которая предусмотрена, причем ее значение должно подтверждаться во время тестирования.

Ниже приведены общие требования к валидации тестирования:

1. При сборе данных:

- сбор данных должен осуществляться одним из двух способов: разработчиком со свидетелем из тестовой лаборатории или путем лабораторных испытаний. Источник энтропии должен содержать интерфейс, который позволяет получать доступ к битам из источника шума и выходным состояниям из компонентов состояний. Этот интерфейс должен позволять получать выходные данные источника шума, то есть, эти результаты не должны использоваться для чего-либо еще хотя бы раз. Интерфейс должен быть доступен во время проверки тестирования, но может быть отключен, если тестирование не проводится;
- данные должны быть получены из источника шума и компонентов состояния, если доступ к таковым имеется, при нормальных условиях эксплуатации;
- данные, полученные из источника энтропии, должны быть не обработанными, а лишь оцифрованными. Может указываться необходимый формат данных для подачи последовательности для тестирования;
- из источника шума должна быть получена необработанная последовательность длины не меньше, чем 1000000 последовательных значений;
- если используется неутвержденный компонент состояния, то длина последовательности для проверки должна быть не меньше 1000000 последовательных значений. Необходимо учитывать, что данные, полученные из источника шума для проверки, могут быть использованы в качестве входных данных для компонентов состояния, чтобы получить выходные данные;
- для выборки значений, состоящих из более чем одного бита, разработчик должен предоставить тестер с упорядоченным ранжированием бит выборки.

2. При тестировании с целью валидации:

- Должно непрерывно осуществляться тестирование работоспособности, причем необходимо проверять, что реализованные тесты обнаруживают отказ с помощью теста счетчика повторов и адаптивного теста долей.
- Тесты должны быть применимы для всех образцов, представленных для тестирования.
- Разработчик должен указать, производит ли источник шума IID данные, либо не IID данные. Это должно использоваться во время определения алгоритма реализации теста валидации. При этом требование полной энтропии будет интерпретироваться, как IID требование. Сущность их состоит в следующем:
- минимальная оценка энтропии, получаемая путем тестирования, должна иметь значение, при котором источник энтропии является валидным. Эта оценка энтропии будет использоваться как минимальная энтропия в выборках;
- источник будет признан таким, что имеет полную энтропию, только после прохождения тестов на IID данные;
- тестировщик будет проверять все данные, а так же рассматривать всю документацию и теоретические обоснования, которые представленные разработчиком.

3. При документировании с целью валидации:

- разработчик должен предоставить документацию, которая описывает работу источника энтропии, а именно: как он работает, как создается энтропия и как получить данные их источника энтропии для их тестирования;

- документация должна быть представлена таким образом, чтобы в лаборатории или поставщик могли выполнять (или повторять) процесс сборки, если это необходимо. Процесс сборки не требует глубоких знаний источника или вмешательств, которые могут изменить поведение источника энтропии;

- документация должна обеспечивать техническую аргументацию, почему источник шума может поддерживать определенный уровень энтропии. Это может быть описание в общих чертах о том, откуда появляется непредсказуемость, а также описание поведения источника шума, чтобы показать его нормальное поведение;

- документация должна содержать описание условий, при которых источник энтропии должен работать корректно (температура, напряжение, системы деятельности). Анализ поведения источника энтропии на границах этих условий должен быть документально оформлен, так же, как и вероятность отказов и повреждений;

- описание тестов работоспособности и основания для этих тестов должны быть включены в документацию. Разработчик должен представить исходный код для любых тестов как альтернативу, либо в дополнение к данным рекомендациям;

- разработчик должен представить описание пространства выходных данных источника шума, в том числе его размер, а также указать размер выборки источника шума, являющегося фиксированной величиной для данного источника шума;

- для источников энтропии, содержащих компоненты состояния, предусматривается их описание, включающее спецификации размера выходного блока компонента состояния.

Общая стратегия тестирования источника энтропии

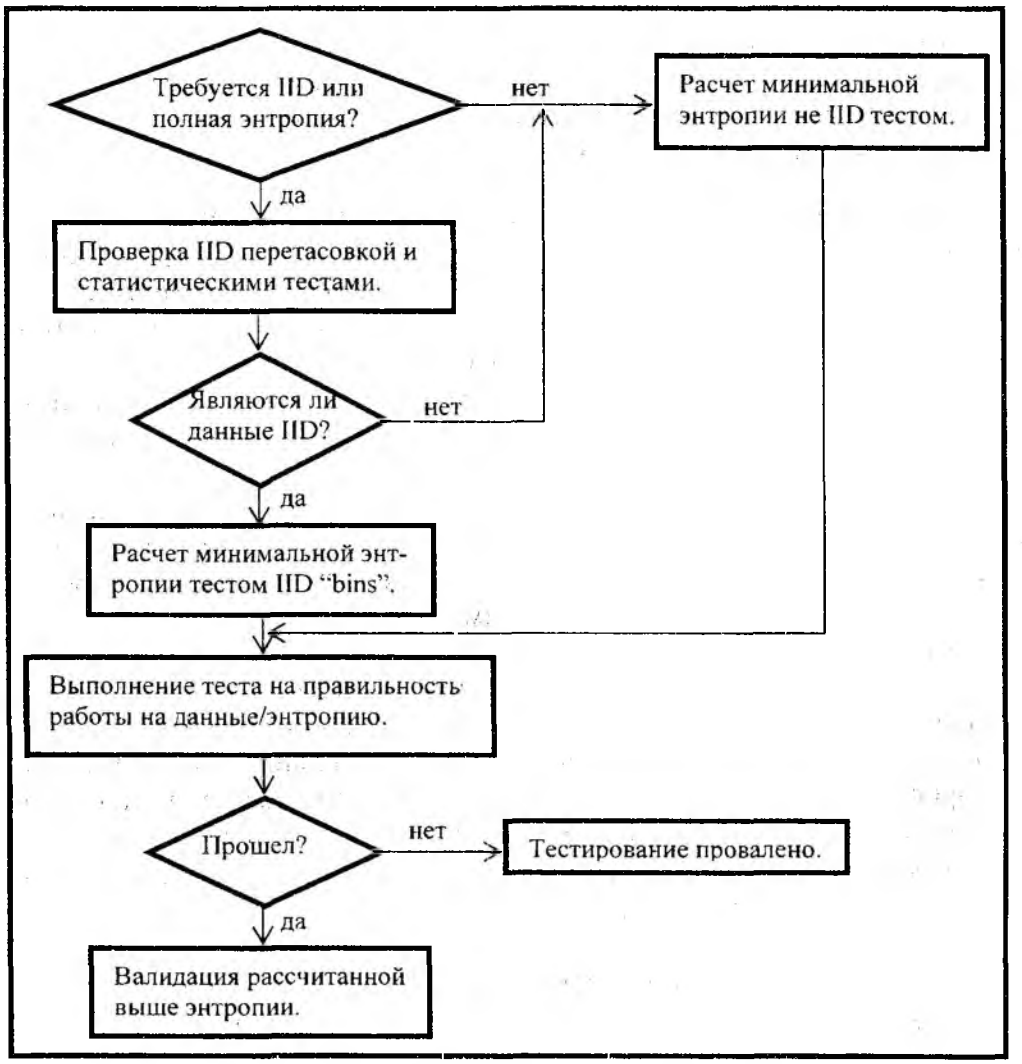
Основным требованием для любого источника энтропии является возможность правильной оценки минимальной энтропии источника. Оценка минимальной энтропии источника энтропии рассчитывается исходя из данных, представленных для тестирования. Как показано на рисунке, первым происходит выяснение, производятся ли IID данные, или нет. Оба результата приводят к оценке минимальной энтропии, но используют при этом для определения оценки энтропии разную информацию. Требования полной энтропии означают, что источник энтропии производит IID данные; любое требование о полной энтропии будет интерпретироваться, как требование IID. Их сущность заключается в следующем:

1. Если требуется, чтобы данные были IID, то полный набор проверок и статистических тестов выполняется для подтверждения того, что данные есть IID.

2. Если результаты тестирования доказали, что данные являются IID, то минимальная энтропия оценивается для источника шума с помощью тестов. Эта оценка должна использоваться для подтверждения минимальной энтропии источника шума, который производит IID данные.

3. Альтернативой, если нет требования IID, или если тесты не поддерживают такого требования, является набор из пяти тестов. Далее, данные пяти тестов приводят к пяти различным оценкам энтропии. При выборе минимальной энтропии выбирается худшая оценка, полученная по тестам, которая и будет в качестве оценки минимальной энтропии источника шума, который не производит IID данные.

4. После получения оценки энтропии данные подвергаются проверке. Эти тесты предназначены для обнаружения существенных сбоев или завышенных значений тестов энтропии. Неспособность пройти такую проверку означает, что источник энтропии не удовлетворяет требования. Если источник энтропии не использует компоненты состояния, но компоненты источника шума прошли тестирование, то считается, что источник энтропии имеет минимальную энтропию.



Стратегия тестирования энтропии

Анализ основных тестов

Одним из основных требований к источнику энтропии является генерирование случайных выходных даных. При этом, для обеспечения достаточной энтропии, подающейся на вход ДГСБ, количество энтропии, образующейся за счет источника шума, должно быть определенным. Далее рассматриваются основные тесты, которые используются при тестировании источников шума, в том числе, когда используются неутвержденные методы, компонент состояний. Эти тесты являются универсальными, но они предназначены в основном для выявления серьезных проблем, которые могут возникать. Тесты предназначены для применения в аккредитованных испытательных лабораториях, но они могут также использоваться и разработчиками источников энтропии.

Тест проверки на независимость и стабильность (перемешивание). Тест позволяет выявить отклонения источника шума или компонент состояния от распределения при независимом и стабильном поведении. Набор данных длины N делится на 10 неперекрывающихся подмножеств равной длины $\left\lfloor \frac{N}{10} \right\rfloor$. Каждое из этих подмножеств тестируется, чтобы определить, что выборки относятся к IID распределению.

В процессе тестирования рассчитываются следующие оценки:

- оценка сжатия – один балл за подмножество данных;

- оценка больших/малых серий – два балла за подмножество данных;
- оценка посещения – один балл за подмножество данных;
- оценка направленных серий – три балла за подмножество данных;
- оценка ковариации – один балл за подмножество данных;
- оценка коллизий – три балла за подмножество данных.

В последующем на каждом из 10 подмножество данных выполняются:

- оценки по оригинальному подмножеству данных, что дает вектор J оценок, где J является количеством баллов для данного теста на каждое подмножество данных;
- следующие шаги повторяются 1000 раз для подмножества данных, которые перемешиваются с помощью генератора псевдослучайных чисел;
– оценки рассчитываются по преобразованным данным, что дает вектор оценок J, где J является количеством баллов для данного теста на одно подмножество данных.
- оценки запоминаются и хранятся в списке;
- для каждого из полученных соответствующих баллов за каждое подмножество данных, оценке оригинальных подмножеств данных присваивается ранг. При этом, если значение оценки оригинального подмножества данных ниже, чем оценка всех преобразованных подмножеств данных, то ранг оригинального подмножества составит 1, и оценки всех других преобразований подмножеств будут иметь более высокий ранг. Если значение оценки оригинального подмножества выше, чем оценка всех преобразованных подмножеств данных, то он будет иметь ранг 1000, а все остальные оценки будут иметь ранг < 1000.

Возможно, что многие преобразованные подмножества данных будут иметь одинаковые оценки. Если оригинал подмножества имеет такую же оценку, как одно или несколько преобразованных подмножеств данных, то она берется ближе к средней. Далее, с учетом отсортированного списка $L[0...1001]$ из оценок преобразованных подмножеств данных и оценки оригинального подмножества S ранг определяется согласно следующему правилу (1):

$$Rank(S) = \begin{cases} \max(j)\ \textit{такое, что}\ L[j] \le S & \textit{если}\ L[500] > S \\ 500 & \textit{если}\ L[500] = S \\ \min(j)\ \textit{такое, что}\ L[j] \ge S & \textit{если}\ L[500] < S \end{cases} \quad (1)$$

Правило (1) гарантирует, что только исключительно высокие или низкие показатели отображаются как аномальные в тесте преобразований (перемешивания).

Далее ранги рассматриваются как p-значения в двухстороннем тесте. С использованием 10 рассматриваемых подмножеств данных формируется набор из $10 \times J$ p-значений для каждого из шести тестов. При этом P-значения объединяются в соответствии со следующими правилами:

- для интервала $50 \le rank \le 950$ вероятность 10 %;
- если 8 или более оригинальных подмножеств данных имеют значение $50 \le rank \le 950$, то источник не проходит испытание;
- если источник шума, подвергнувшись тестированию, не проходит испытание на независимость и стабильность, то он не считается IID. Далее тестирование источника шума необходимо испытывать по альтернативному пути;
- если неутвержденный компонент состояния проходит тест, источник энтропии не проходит валидацию.

К примеру, если S = 20, L[500] = 22, L[299] = 19, L[300] = 20, L[301] = 20 и L[302] = 22, тогда Rank(S) = 301. Если S = 20, L[500] = 18, L[599] = 19, L[600] = 20 и L[601] = 20, тогда Rank(S) = 600.

Тест по оценке сжатия. Универсальные алгоритмы сжатия эффективно применяются для удаления избыточности в строке символов, особенно при наличии повторяющихся в последовательностях символов. Оценка сжатия подмножества данных, оригинальных или

преобразованных, изменяет длину подмножества. Баллы за сжатие вычисляются следующим образом:

- выборка в подгруппе данных кодируется в виде строки символов, содержащих список значений, разделенных запятыми;
- символьная строка обрабатывается при помощи, например, [BZ2] алгоритма сжатия;
- определяется длина сжатой строки в байтах.

Оценка больших/малых серий. В процессе этой оценки вычисляется среднее значение подмножеств данных и все значения, не равные ему либо большие, либо ниже среднего. Длинные серии, которые выше либо ниже серии, должны появляться сравнительно редко, если выходные данные являются независимыми и стабильными. Кроме того, число серий выборок, которые находятся выше или ниже среднего, являются мерой, показывающей, является ли данная выборка зависимой от соседних и устойчиво ли распределение.

Этот тест применим для любого подмножества данных, выборки берутся в числовом или порядковом значении. Тем не менее, тест должен использоваться, даже если данные не являются числовым или порядковым номером.

Каждое подмножество данных (оригинальное или преобразованное), в дальнейшем используется для получения временного подмножества данных, состоящих только из значений -1 и +1. Значениям, которые меньше среднего, присваиваются -1 во временном подмножестве данных, а тем, что больше среднего, присваивается +1. Если значение равно среднему, то оно упускается из временного подмножества. Самая длинная серия и количество серий во временном подмножестве отмечаются, ей дается на 2 балла меньше для каждого из оригинального или преобразованного подмножества данных.

Баллы рассчитываются следующим образом:

1) подмножества данных используются для вычисления среднего значения, если данное подмножество является двоичным, то среднее будет 0.5;

2) временное подмножество строится следующим образом – для каждого из оригинальных и преобразованных подмножеств данных:

- если элемент больше, чем среднее, то добавляем +1 к временному подмножеству,
- если элемент меньше, чем среднее, то добавляем -1 к временному подмножеству,
- если элемент такой же, как и средний, то опускаем его во временном подмножестве;

3) определяется самая длинная серия -1 или +1 во временном подмножестве данных и длинна этой серии будет первой оценкой;

4) число серий из -1 и +1 во временном подмножестве используется как вторая оценка.

К примеру, пусть подмножество данных состоит из 7 значений {5, 15, 12, 1, 13, 9, 4}. Средним значением для этого подмножества будет 9. В таком случае временным подмножеством будет {-1, +1, +1, -1, +1, -1}. Получим серии (-1), (+1, +1), (-1), (+1), (-1). Таким образом, длина наибольшей серии составляет 2 (первая оценка), а количество серий составляет 5 (вторая оценка).

Оценка «посещений». Определяется, когда имеется последовательность высоких или низких значений оцено(кластеров). Это факт указывает на то, что распределение может не быть стабильным и независимым. Потому, как высокие, так и низкие баллы «посещений» представляют интерес. Высокие показатели представляют собой оценки, которые принимают значение из области высокого/низкого значения. Низкие указывают на некоторый процесс, который предотвращает необычно высокие/низкие значения от кластеризации вместе.

Оценки «посещения» имеют смысл, если среднее значение из выборки имеет смысл, например, если среднее может быть вычислено по набору значений данных. В большинстве случаев, в которых выборка значений является счетчиком или оцифрованным значением, среднее значение выборки данных может быть вычислено. Тем не менее, тестирование должно выполняться, если даже среднее не может быть вычислено.

Оценка «посещений» является мерой того, насколько текущая сумма значений отсчетов отклоняется от ожидаемого значения в каждой точке множества данных. Если подмножество данных, $s_0, s_1, s_2,...$, а среднее из выборки значений является μ, то «посещ ение» в позиции i будет $s_0 + s_1 + ... + s_i - i \times \mu$. Возвращаемая оценка является абсолютным максимальным значением любого «посещения» в подмножестве данных.

В этом случае оценка вычисляется следующим образом:

- для $j = 1 \, до \left\lfloor \frac{N}{10} \right\rfloor$:

d_j _ абсолютное значение (сумма первых j выборок – $j \times \mu$);

- возвращаемая оценка является наибольшим d_j значением.

К примеру, если подмножество данных {2, 15, 4, 10, 9}, тогда $\mu = 8$. $d_1 = |2 - 8| = 6$; $d_2 = |(2+15) - (2 \times 8)| = 1$; $d_3 = |(2+15+4) - (3 \times 8)| = 3$; $d_4 = |(2+15+4+10) - (4 \times 8)| = 1$; $d_5 = |(2+15+4+10+9) - (5 \times 8)| = 0$. Получим, что наибольшее значение $d_j = 6$.

Оценка направленных серий. Многие физические процессы лучше всего понимаются в терминах их производных. Если в первой производной знаки чередуются между выборками, то в последовательности будет присутствовать большое количество коротких направленных серий. Каждое подмножество данных (оригинал или преобразование) используется для создания временного подмножества данных, такого, что один элемент короче, чем оригинал подмножества данных. Временное подмножество содержит указания о том, что первое значение в каждой паре элементов меньше, равно или больше второго элемента в паре. Если первый элемент меньше, то добавляется +1 к временному подмножеству, если элементы равно, то добавляется 0 к временному подмножеству, если первый элемент больше, то добавляется -1 к временному подмножеству. Оценки рассчитываются по алгоритму, описанному ниже, где s_i – i-й элемент подмножества данных, а вес Хэмминга определяется как число единиц в последовательности.

Временное подмножество данных формируется из подмножества данных:

1) Если источник является недвоичным, то :

для $i = 0$ до (длина оригинального подмножества данных) – 2:

- если $s_i < s_{i+1}$, тогда $temp_i = 1$;

- если $s_i > s_{i+1}$, тогда $temp_i = -1$;

- в противном случае $temp_i = 0$.

2) Если источник последовательности является двоичным, то подмножество битов объединяется в байты, и выполняется:

- для $i = 0$ до (длина оригинального подмножества данных)/8 – 1:

$$W_i = вес_Хэмминга(s_i,...,s_{i+7}) \; ;$$

- для $i = 0$ до (длина последовательности W) – 2:

-если $W_i < W_{i+1}$,*тогда* $temp_i = 1$;

- если $W_i > W_{i+1}$,*тогда* $temp_i = -1$;

- в противном случае $temp_i = 0$.

Затем выполняется расчет баллов по временному подмножеству данных.

Пример с недвоичным подмножеством данных {2, 2, 2, 5, 7, 7, 9, 3, 1, 4, 4}. Получим временное подмножество {0, 0, +1, +1, 0, +1, -1, -1, +1, 0}. Таким образом получаем три серии: (+1, +1, 0, +1), (-1, -1), (+1, 0), следовательно, первая оценка равна 3. Длина самой длинной серии равна 4 (вторая оценка). С учетом того, что во временном подмножестве

четыре “+1” и две “-1”, то третья оценка равна 4.

Пример с двоичным подмножеством данных. Биты объединяются в байты и последовательность имеет вид {A3, 57, 3F, 42, BD}. Вес Хэмминга соответственно {4, 5, 6, 2, 6}. Временным подмножеством будет {+1, +1, -1, +1}. Получим три серии (+1, +1), (-1), (+1), и первая оценка равна 3. Самая длинная серия равна 2 (вторая оценка). Мы имеем три значения “+1” и одно значение “-1”, значит третья оценка равна 3.

Оценка ковариации. Есть случаи, когда источник шума не связан с IID зависимостями между соседними выборками. Тест на независимость Хи – квадрат и оценка сжатия являются эффективными для обнаружения повторяющихся значений. Но он не будет выявлять связи между числовыми значениями последовательных выборок.

Поведение ковариационной оценки рассмотрим для двумерного нормального распределения. В этом случае любая линейная зависимость между последовательными парами значений повлияет на этот результат, и поэтому ковариация будет различной для исходного подмножества данных и для преобразованного подмножества данных, если такая связь существует. Ковариация двух переменных определяется как ожидаемое значение их результата, за вычетом результата их ожидаемого значения. Ковариация является общим статистическим показателем наличия линейной зависимости между двумя переменными. Ковариационный «балл» вычисляется для каждой пары подмножества данных S, но таким образом, что s_0 является парой для s_1, s_1 для s_2, и т.д.

Оценка вычисляется следующим образом:

- счетчик $count = 0$;
- вычисляется из $s_0, s_1, ...s_{N/10-1}$ μ – среднее значение;
- для $i = 0$ *до* $\left\lfloor \frac{N}{10} \right\rfloor$:

$count = count + (s_i - \mu)(s_{i-1} - \mu)$.

Баллы вычисляются как $\frac{count}{\left\lfloor \frac{N}{10} \right\rfloor - 1}$.

К примеру, если подмножество данных будет {15, 2, 6, 10, 12}, тогда $\mu = 9$. Для $i = 1$, $count = (2-9)(15-9) = -42$;

$i = 2$, $count = -42 + (6-9)(2-9) = -21$;

$i = 3$, $count = -21 + (10-9)(6-9) = -24$;

$i = 4$, $count = -24 + (12-9)(10-9) = -21$.

Тогда оценка будет равна $(-21/4) = -5$.

Оценка коллизий. Энтропия зависит от того, сколько раз от источника получены случайные данные перед первым их повторением, т.е. коллизией. От источника IID можно ожидать одинаковое поведение относительно коллизии для преобразованных данных. При этом, оценка коллизий сводится к определению появления выборок из двух повторяющихся комбинаций. Однако возможны различные коллизии, в том числе и многократные.

Начиная с начала подмножества данных рассматриваются выборки, пока не будет найдено повторение. Число исследованных выборок фиксируется, и новый поиск начинается со следующего элемента подмножества. Это продолжается, пока не будет проверено все подмножество данных. Набор чисел из исследованных образцов затем используется для получения трех оценок: наименьшего числа, среднего числа и наибольшего числа.

В случае, если данные двоичные, подмножество, состоящее из двоичных данных, преобразуется в последовательность 8-битных байт таким образом, что каждая последовательность из 8 бит в оригинальном подмножестве становится одним байтом в модифицирован-

ном подмножестве данных. Затем измененное подмножество подвергают тестированию для вычисления баллов коллизии. Необходимо заметить, что оригинальное и преобразованное подмножества модифицируются отдельно. Необходимо отметить, что, что длина модифицированного подмножества равна $\left\lfloor \frac{N}{80} \right\rfloor$, в то время как длина исходного двоичного и недвоичного подмножества равны $\left\lfloor \frac{N}{10} \right\rfloor$.

Баллы рассчитываются следующим образом:

- *Counts* – это список счетчиков, который используется для нахождения повторения(коллизий) в выборках, изначально пуст – $Counts = 0$;

- $pos = 0$.

- пока $pos <$ (длина подмножества данных):

- находим наименьшее j такое, что $s_{pos} \ldots s_{pos+j}$ содержат одно повторяющееся значение выборки;

- если такого j не существует – прерываем цикл;

- добавить j в список *Counts*;

– $pos = pos + j + 1$.

В заключение вернуть следующие значения как оценок: минимальное значение в списке *Counts*, среднее значение в списке *Counts*, максимальное значение в списке *Counts*.

К примеру, пусть подмножество данных будет {2, 1, 1, 2, 0, 1, 0, 1, 1, 2} длиной, равной 10. Первое значение "2" находится на нулевой позиции, тогда первая коллизия происходит, когда $j = 2$ (единицы на первой и второй позиции). Не обращая внимания на первые три значения подмножества, необходимо рассмотреть коллизии в подмножестве {2, 0, 1, 0, 1, 1, 2}. Они возникают при $j = 3$ (нули на первой и третьей позиции). Третьим подмножеством будет {1, 1, 2}, в котором коллизия будет при $j = 1$ (нулевая и первая позиции равны 1). Последнее подмножество {2}, для которого нет коллизий. Список счетчиков для этого теста будет $j = 2, j = 3, j = 1$. Первая оценка будет 1 (минимальное значение счетчика), вторая оценка 2 (среднее значение счетчиков), третья оценка 3 (максимальное значение счетчика).

Выводы

В стандарте NIST SP 800-90b приведены требования, которые выдвигаются в процессе создания и применения источника энтропии, обоснование этих требований и методики тестирования. При этом тестирование направлено на проверку корректной работоспособности и удовлетворения критериям источника энтропии, минимальной энтропии, полного источника энтропии, генератора шума, а также проверки компонентов состояния.

На сегодняшний день в Украине нет аналогичного национального стандарта, который содержал бы требования и методики тестирования источников энтропии, источников шума, определения минимальной энтропии, полной энтропии источника и компонентов состояния. Можно сделать вывод, что данный стандарт является мощным инструментом в создании и тестировании генераторов случайных последовательностей и является хорошим кандидатом на гармонизацию с последующим принятием его в качестве государственного стандарта Украины.

Список литературы: 1. *Federal* Information Processing Standards Publication (FIPS PUB) 140-2. Security requirements for cryptographic modules. NIST, 1999. 2. *National* Institute of Standards and Technology, FIPS 140-3 (DRAFT), Security for cryptographis modules: [Електронний ресурс]. Режим доступу: http://www. nist. gov/cmvp 3. *ISO/IEC* FCD 19790: Information technology- Security requirements for cryptographicmodules. Proect: 1.27.40. 4. *NIST SP 800-22.* A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications: [Електронний ресурс]. April 2000. 5. *NIST* DRAFT Special Publication 800-90B. Recommendation for the Entropy Sources Used for Random Bit Generation / E. Barker, J. Kelse. 2012.

Харьковский национальный университет радиоэлектроники

Поступила в редколлегию 15.09.2012

УДК 681.3.06

Е.В. КОТУХ

УНИВЕРСАЛЬНОЕ ХЕШИРОВАНИЕ С ОГРАНИЧЕНИЕМ ФУНКЦИОНАЛЬНОГО ПОЛЯ АЛГЕБРАИЧЕСКИХ КРИВЫХ

Построение доказуемо стойкой аутентификации на основе универсального хеширования по рациональным функциям алгебраических кривых позволяет разрешить основное противоречие между затратами ключа на аутентификацию, которые определяются размером сообщений и значением вероятности коллизии (обмана), которая определяется пространством хеш кодов. Наилучший результат универсального хеширования достигается на максимальных кривых, число точек которых лежит на границе Хассе – Вейля [1 – 3]. Вычислительная и алгоритмическая сложность хеширования зависит от размерности проективного пространства представления кривой. Проективное пространство алгебраической кривой определяется базисом линейного векторного пространства Римана – Роха для функционального поля кривой. Существуют три замечательных семейства максимальных кривых, которые связываются с Дэлигнэ – Лустига (Deligne – Lusztig) многообразием размерности $\dim = 1$. Кривая Дэлигнэ – Лустига ассоциируется с проективной специальной линейной группой (кривые Эрмита и кривые которые покрываются кривой Эрмита), с группой Сузуки (Suzuki) $Sz(q)$ (кривые Судзуки) и Ри (Ree) группой $R(q)$ [4]. В работах [3, 5] представлено универсальное хеширование по кривой Судзуки над конечным полем характеристики 2 с нечетной степенью расширения. Кривая Судзуки имеет определение в проективном пространстве P^4 и требует в два раза больше хеш вычислений по сравнению с хешированием по кривым Эрмита в P^2 для квадратичного поля. Актуальной задачей является разработка метода универсального хеширования по алгебраическим кривым с уменьшенной сложностью вычислений.

В данной работе предлагается метод универсального хеширования с ограничением функционального поля алгебраических кривых. С этой целью в разд. 1 приводятся определение и свойства универсального хеширования по рациональным функциям алгебраических кривых. В разд. 2 представлен метод универсального хеширования с ограничением функционального поля алгебраических кривых, коллизионные оценки и оценки сложности хеширования.

1. Определения и свойства универсального хеширования по рациональным функциям алгебраических кривых

Универсальное хеширование по рациональным функциям алгебраической кривой представляется определением 1 [6].

Определение 1. Пусть

χ – абсолютно неразложимая, несингулярная проективная кривая над полем F_q;

$P_1, P_2, ..., P_n$ – точки кривой χ;

P_∞ – точка на бесконечности или особая точка кривой χ;

$f_i \in F_q(\chi) \setminus \{0\}$ – рациональные функции поля рациональных функций кривой χ;

$div_\infty(f_i) = \rho_i$ значение дивизора или порядок полюса рациональной функции f_i в точке P_∞;

$f_i(P_j)$ – значение рациональной функции в точке P_j.

Хеш функция $h_{P_j}(m) \in F_q$ для сообщения $m = (m_1, ..., m_k)$, $m_i \in F_q$ в точке P_j определяется выражением

$$h_{P_j}(m) = \sum_{i=1}^{k} f_i(P_j) m_i, \qquad (1)$$

где $f_i \in F_q(\chi)$ с упорядоченными порядками полюсов $0 < \rho_1 < \rho_2 < ... < \rho_k$ ..

Свойства универсального хеширования по рациональным функциям алгебраических кривых определяются утверждением 1.

Утверждение 1 [6]. Хеш функция $h_{P_j}(m)$ определяет универсальный хеш класс $\varepsilon - U(N, q^k, q)$, где N – число точек алгебраической кривой, q^k – объём пространства сообщений, q – объём пространства хеш кодов и вероятность коллизии определяется выражением

$$\varepsilon = \rho_k / N, \tag{2}$$

где ρ_k – значение полюса рациональной функций f_k.

Замечание 1.

1. Параметры универсального хеш класса $\varepsilon - U(N, q^k, q)$ на основе хеширования по рациональным функциям определяются свойствами алгебраической кривой. Подгруппа Вейерштрасса $H(P_\infty) = \{\rho_0 = 0 < \rho_1 < ...\}$ определяется полюсами рациональных функций в особой точке кривой и рациональные функции упорядоченные по значениям полюсов образуют векторное линейное пространство размерности $\dim(L(G)) = \nu_\ell := \#\{(i, j) \in \mathrm{N}^2 : \rho_i + \rho_j = \rho_{\ell+1}\}$.

2. Ключевой параметр хеш функции $h_{P_j}(m)$ определяется вычислением в точке алгебраической кривой.

3. Асимптотическая граница вероятности коллизии для $\varepsilon - U(N, q^k, q)$ хеш класса построенного по рациональным функциям плоских алгебраических кривых рода g над большим алфавитом и фиксированных $k \le g$ и q, имеет вид [8]

$$1 - \frac{q^k(q-1)}{(q^k-1)q} \le P_{кол} \le \varepsilon \le \frac{\sqrt{2k}}{q} + \frac{3\sqrt{2k}}{2q\sqrt{q}}.$$

Верхняя асимптотическая граница вероятности коллизии лучше асимптотической границы для хеширования по алгебраическим кодам и уточняет границу для алгеброгеометрических кодов.

Наилучшие результаты универсального хеширования по соотношению затрат на поле вычислений, размер ключевых данных при фиксированном значении вероятности коллизии достигаются на алгебраических кривых с наибольшим отношением рода g к числу точек.

Главный результат для максимальных кривых Дэлигнэ – Лустига представлен в теоремах 1 и 2.

Теорема 1 [7]. Пусть C кривая над F_q рода g и удовлетворяются следующие условия

1. $g > (\sqrt{q} - 1)^2 / 4$

2. $\#C(F_q) = q + 2g\sqrt{q} + 1$, (то есть C является максимальной над F_q).

Тогда X является F_q изоморфной кривой Эрмита над F_q и её род $g = \sqrt{q}(\sqrt{q} - 1)/2$.

Теорема 2 [8]. Для положительного целого s заданы $q = 2q_0^2$ и $q_0 = 2^s$. Пусть X кривая над F_q рода g и удовлетворяются следующие условия:

1. $g = q_0(q - 1)$;

2. $\#X(F_q) = q^2 + 1$.

Тогда X является F_q изоморфной кривой Дэлигнэ-Лустига ассоциированной с группой Судзуки $Sz(q)$.

Размерность функционального поля кривой Эрмита определяется леммой 1.

Лемма 1. Пусть P есть рациональная точка на кривой Эрмита над полем F_q, $q = l^2$. Тогда подгруппа Вейерштрасса $H(P) = \langle l, l+1 \rangle$. Кривая Эрмита является максимальной и определяется линейной серией размерности $\dim = 2$.

Результаты по максимальным плоским кривым в конечном поле F_q, $q = l^2$ представлены в таблице.

Уравнение кривой $C(F_{l^2})$	Значение рода кривой	Полюса рациональных функций	Значение подгруппы Вейерштрасса
$+y = x^{l+1}$	$g_1 = l(l-1)/2$	$(x)_\infty = l, (y)_\infty = l+1$	$\langle l, l+1 \rangle$
$+y = x^{(l+1)/2}$, l нечетное	$g_2 = l(l-1)^2/4$	$(x)_\infty = l, (y)_\infty = (l+1)/2$	$\langle (l+1)/2, l \rangle$
$y^{l/2^t} = x^{l+1}, l = 2^t$	$g_2' = l(l-2)/4$	$(x)_\infty = l/2, (y)_\infty = l+1$	$\langle l/2, l+1 \rangle$
$+y = x^{(l+1)/3}, l \equiv 2 (\mathrm{mod}\, 3)$	$g_3' = (l^2-3l+2)/6$	$(x)_\infty = (l+1)/3, (y)_\infty = l$	$\langle (l+1)/3, l \rangle$
$y^{3^t} = \omega x^{l+1}, l = 3^t,$ $\in F_{l^2}, \omega^{l-1} = -1$	$g_3'' = l(l-3)/6$	$(x)_\infty = l/3, (y)_\infty = l+1$	$\langle l/3, l+1 \rangle$ $\langle 2(l+1)/3, l, l+1 \rangle$
$^{l+1)/3} + x^{2(l+1)/3} + y^{l+1} = 0,$ $\equiv 2 (\mathrm{mod}\, 3)$	$g_3 = (l^2-l+4)/6$	$(x)_\infty = l+1, (y)_\infty = (l+1)/3,$ $(v)_\infty = l$	$\langle 2(l+1)/3, l, l+1 \rangle$
$x^{(l-1)/3} - yx^{2(l-1)/3} + y^l = 0,$ $\in F_{l^2}, \omega^{l-1} = -1, l \equiv 2 (\mathrm{mod}\, 3)$	$g_3' = l(l-1)/6$	$(x)_\infty = l, (y)_\infty = (l-1)/3,$ $(v)_\infty = l+1$	$\langle 2(l-1)/3, l, l+1 \rangle$
$^l + y = (\sum_i^t x^{l/3^i})^2, l = 3^t$	$g_3' = l(l-1)/6,$	$(x)_\infty = l, (y)_\infty = 2l/3,$ $(v)_\infty = l+1$	$\langle 2l/3, l, l+1 \rangle$
$^{(l+1)/3} y^{(l+1)/3} + y^{2(l+1)/3} + x^{(l+1)/3} = 0$	$g_3 = (l^2-l+4)/6$		$\langle 2(l+1)/3, l, l+1 \rangle$

Замечание 2.

1. Алгебраические кривые $y^l + y = x^{l+1}$, $y^l + y = x^{(l+1)/2}$ и $\sum_{i=1}^{t} y^{l/2^t} = x^{l+1}, l = 2^t$ являются максимальными кривыми первого и второго рода, имеют подгруппу Вейерштрасса $H(P_\infty) = \langle \rho_1, \rho_2 \rangle$ размерности $\dim = 2$ и функциональное поле определяется функциями вида $\{x^i \cdot y^j\}$.

2. Алгебраические кривые $y^l + y = x^{(l+1)/3}, l \equiv 2 (\mathrm{mod}\, 3)$ и $\sum_{i=0}^{t-1} y^{3^i} = \omega x^{l+1}, l = 3^t, \omega \in F_{l^2}, \omega^{l-1} = -1$ являются максимальными кривыми третьего рода, имеют подгруппу Вейерштрасса $H(P_\infty) = \langle \rho_1, \rho_2 \rangle$ размерности $\dim = 2$ и функциональное поле определяется функциями вида $\{x^i \cdot y^j\}$.

3. Максимальные кривые вида:
   - $x^{(l+1)/3} + x^{2(l+1)/3} + y^{l+1} = 0, l \equiv 2 (\mathrm{mod}\, 3)$,
   - $\omega x^{(l-1)/3} - yx^{2(l-1)/3} + y^l = 0, \omega \in F_{l^2}, \omega^{l-1} = -1, l \equiv 2 (\mathrm{mod}\, 3)$,
   - $y^l + y = (\sum_i^t x^{l/3^i})^2, l = 3^t$

имеют подгруппу Вейерштрасса $H(P_\infty)$ размерности $\dim = 3$ и функциональное поле определяется рациональными функциями вида $\{x^i \cdot y^j \cdot v^l\}$.

Кривая Дэлигнэ-Лустига, ассоциированная с группой Судзуки, определяется полной линейной серией $D = |(q + 2q_0 + 1)P_0|$ размерности $\dim = 4$ и степени $q + 2q_0 + 1$, которая выводится из энумератора зета функции. Кривая Судзуки имеет отображение на проективное пространство P^4 и подгруппу Вейерштрасса $H(P) = \langle q, q + q_0, q + 2q_0, q + 2q_0 + 1 \rangle$, $P \in X(F_q)$ [8].

Таким образом, практические алгоритмы вычисления хеш функций определяются многопараметрическими схемами вычисления по базису векторного пространства функционального поля алгебраической кривой. Сложность вычисления хеш функции определяется размерностью базисного пространства рациональных функций функционального поля. Так хеширование по кривым Судзуки над четырех мерным базисом в случае оптимизации вычислений по схеме Горнера в два раза сложнее чем хеширование по кривой Эрмита над двумерным базисом [3]. Ограничение функционального поля кривых приводит к уменьшению сложности вычислений.

2. Метод универсального хеширования с ограничением функционального поля алгебраических кривых

Определение 2. Пусть χ – абсолютно неразложимая, несингулярная проективная кривая над полем F_q с точками $P_1, P_2, ..., P_n$, особой точкой (точкой на бесконечности) P_∞ и функциональным полем $F_q(\chi) \setminus \{0\}$. Хеш функция $h_{P_j}(m)$ над конечным полем F_q для сообщения $m = (m_1, ..., m_k)$, $m_i \in F_q$ в точке P_j кривой χ определяется выражением

$$h_{P_j}(m) = \sum_{i=1}^{k} f_i(P_j) m_i, \qquad (3)$$

где $f_i \in G_q(\chi)$ – рациональные функции подмножества функционального поля $G_q(\chi) \in F_q(\chi) \setminus \{0\}$ с упорядоченными порядками полюсов $0 < \rho_1 < \rho_2 < ... < \rho_k$, $\rho_i = div_\infty(f_i)$ значение дивизора или порядок полюса рациональной функции в точке P_∞.

Замечание 3.

1. Метод универсального хеширования с ограничением функционального поля алгебраических кривых предусматривает следующий порядок действий:

- задание алгебраической кривой χ над полем F_q, вычисление точек кривой $P_1, P_2, ..., P_n$;

- вычисление функционального поля $F_q(\chi) \setminus \{0\}$ ассоциированного с алгебраической кривой χ;

- ограничение функционального поля подмножеством рациональных функций $f_i \in G_q(\chi)$ с упорядоченными порядками полюсов;

- построение алгоритма вычисления хеш функции на ограниченном подмножестве рациональных функций.

2. Ограничение функционального поля алгебраических кривых определяется решением задачи минимизации сложности вычислений при хешировании заданного числа слов данных и заданной вероятности коллизии за счет оптимизации выбора базисных функций и размерности функционального пространства.

Рассмотрим решение задачи оптимизации выбора базисных функций и размерности функционального пространства для хеширования по полю рациональных функций кривой Судзуки.

Утверждение 2. Универсальное хеширование по кривой Судзуки над F_q, $q = 2q_0^2$, $q_0 = 2^s$ с ограничением функционального поля в базисе линейного пространства $L(mP_\infty) = L(\rho_l P_\infty)$, $\rho_l \le m \le \rho_{l+1}$ рациональных функций $\{x^i \cdot y^j : iq + j(q + 2q_0) \le m\}$ определяет $U(q^2, q^k, q)$ семейство хеш функций с вероятностью коллизии

$$\varepsilon = k/(2qq_0) + s/q - s(s-1)/(4qq_0), \text{ если } \rho_k \le 2q_0(q-1) \tag{4}$$

и сложностью хеширования

$$N_{xy} = k + s \tag{5}$$

где k – число слов данных, $s = \left\lceil (2k + 1/4)^{1/2} - 1/2 \right\rceil$, $\lceil \cdot \rceil$ округление к большему целому числу.

Доказательство. Кривые Судзуки S являются F_q изоморфными плоской кривой $Y^q Z^{q_0} - YZ^{q+q_0-1} = X^{q+q_0} - X^{q_0+1} Z^{q+q_0-1}$, где $q = 2q_0^2$ и $q_0 = 2^s$. Род кривой $g = q_0(q-1)$ и число F_q рациональных точек равно $q^2 + 1$. Точками кривой являются особая точка на бесконечности $P_0 = (0:1:0)$ кратности q_0 и рациональные точки $P_{a,b} = (a:b:1)$, где $a, b \in F_{q^2}$ и $b^q - b = a^{q_0}(a^q - a)$. Подгруппа Вейерштрасса функционального поля кривой содержит подгруппу $H(P_\infty) = \langle q, q + q_0, q + 2q_0, q + 2q_0 + 1 \rangle$. Базис пространства $L(\rho_\ell P_0)$, задается функциями вида $\{w^j \cdot v^i \cdot y^t \cdot x^r : i(q + 2q_0) + j(q + 2q_0 + 1) + t(q + q_0) + r \cdot q \le \rho_\ell\}$, что следует из подгруппы Вейерштрасса $H(P_0)$ представленной порядками полюсов функций $x = X/Z$, $y = Y/Z$, $v = x^{2q_0+1} + y^{2q_0}$, $w := xy^{2q_0} + x^{2q+2q_0} + y^{2q}$. Порядки полюсов равны $div_\infty(x) = qP_0$, $div_\infty(y) = (q + q_0)P_0$, $div_\infty(v) = (q + 2q_0)P_0$, $div_\infty(w) = (q + 2q_0 + 1)P_0$.

Хеш функция $h_{x,y}(m) \in F_q$, для сообщения m по рациональным функциям кривой Судзуки в точке x, y определяется выражением

$$h_{x,y}(m) = \sum m_{i,j,t,r} \cdot w^j \cdot v^i \cdot y^t \cdot x^r,$$

где $m_{i,j,t,r} \in F_q$ – слова сообщения m, $i \ge 0$, $0 \le j \le 2q_0 - 1$, $0 \le t \le 1$, $0 \le r \le q_0$, $i(q + 2q_0) + j(q + 2q_0 + 1) + t(q + q_0) + rq \le \rho_k$, $x = X/Z$, $y = Y/Z$, $v = x^{2q_0+1} + y^{2q_0}$, $w := xy^{2q_0} + x^{2q+2q_0} + y^{2q}$, ρ_k полюс подгруппы Вейерштрасса $H(P_\infty)$ [3].

Кривая Судзуки при отображении на проективное пространство P^4 представляется точками вида $P_{(a,b)} := (1:a:b:f(a,b):af(a,b) + b^2) \cup \pi(P_0) = (0:0:0:0:1)$, где $a, b \in F_q$ и $f(a,b) := a^{2q_0+1} + b^{2q_0}$.

Пусть $L(\rho_\ell P_0)$ определяется базисными функциями $\{x^i \cdot y^j : i \cdot q + j(q + q_0) \le \rho_\ell\}$. Для линейного пространства $L(\rho_\ell P_0)$, построенного по полному функциональному базису кривой Судзуки, имеем

$$\rho_{g+1} = 2g = 2q_0(q-1) = q + (2q_0 - 2)(q + q_0), \tag{6}$$

где g – род кривой.

Из условия (6) следует, что $i + j \le 2q_0 - 1$ и получим вычисление хеш кода по двух параметрической схеме Горнера эквивалентной схеме вычисления по кривой Эрмита

$$h_{x,y}(m) = \sum_{j=0}^{s} y^j \cdot \sum_{i=0}^{s-j} m_{i,j} \cdot x^i,$$

где $s=\left|(2k+1/4)^{1/2}-1/2\right|$ – параметр от k слов данных (лемма 1 [9]).

Множество ключей определяется пространством значений x, y. Универсальное хеширование выполняется по базису $\{x^i \cdot y^j : i \cdot q + j(q+q_0) \le \rho_k\}$. Таким образом получим $U(q^2, q^k, q)$ семейство хеш функций и вероятность коллизии $\varepsilon = \rho_k / N = \rho_k / q^2$. Нетрудно показать (см. лемма 1 [9]), что $j = k - s(s-1)/2$, $i = s - j$ и $s=\left|(2k+1/4)^{1/2}-1/2\right|$ при $\rho_k \le \rho_{g+1}$, где $\lceil\cdot\rceil$ округление к большему целому числу. Подставим соотношения i, j, s в выражение для ε и получим

$$\varepsilon = \rho_k / q^2 = (i \cdot q + j(q+q_0))/q^2 = k/(2qq_0) + s/q - s(s-1)/(4qq_0). \qquad (7)$$

Оценка сложности универсального хеширования по базису x, y следует из оценки сложности универсального хеширования по двух параметрической схеме Горнера для кривой Эрмита (предложение 2 [9])

$$N_{xy} = k + s, \text{ если } \rho_k \le \rho_{g+1}, \qquad (8)$$

где $s=\left|(2k+1/4)^{1/2}-1/2\right|$.

Следствие 1. Асимптотическая оценка вероятности коллизии универсального хеширования по кривой Судзуки имеет вид

$$\varepsilon \approx 1/q_0, \text{ если } k \le (2q_0-1)(q_0+1). \qquad (9)$$

Доказательство. Число хешируемых слов данных определяется числом полюсов подгруппы Вейерштрасса $\rho_k \le \rho_{g+1}$. Число пар $i+j \le 2q_0-1$ определяется как сумма арифметического ряда вида

$$k_{xy} = 1+2+\ldots+(2q_0-1)+2q_0-1 = (2q_0-1)(q_0+1). \qquad (10)$$

Коэффициенты i, j в (7) для $k_{xy} = (2q_0-1)(q_0+1)$ слов данных имеют значения $i=1$, $j = 2q_0-2$. Окончательно получим

$$\varepsilon = \rho_k / q^2 = (q + (2q_0-2)(q+q_0))/q^2 = (2q_0-1)/q + 1 - 2/q^2 \approx 1/q_0. \quad \Diamond$$

Замечание 2.

1. Оценка вероятности коллизии хеширования по кривой Судзуки при полном функциональном базисе для $k_{xy} = (2q_0-1)(q_0+1)$ имеет вид (см. замечание 1 [3])

$$\varepsilon = (3k_{xy})^{1/3}/q^2 = ((2q_0-1)(q_0+1))^{1/3}(q+q_0)/q^2 \approx 1.6/(q_0 q_0^{1/3}). \qquad (11)$$

Сравнение с асимптотической оценкой (9) показывает, что хеширование по двух параметрическому базису кривой Судзуки в $q_0^{1/3}$ раз проигрывает по вероятности коллизии.

2. Сложность универсального хеширования по кривым Судзуки в полном функциональном базисе определяется выражением (см. предложение [5])

$$N_{xyvw} = k + s^3/3 + s^2/2 + 2s - 1, \text{ если } s \le q_0, \qquad (12)$$

$$N_{xyvw} = k + q_0^3/3 + q_0^2/2 + (s-q_0)(2q_0-1) + 2s - 1, \text{ если } s > q_0, \qquad (13)$$

где $s = (3k)^{1/3}$.

Асимптотика оценки сложности универсального хеширования по кривым Судзуки следует из (12). При $s = (3k)^{1/3}$ и $s \le q_0$ получим

$$N_{xyvw} = 2k + (3k)^{2/3}/2 + 2(3k)^{1/3} - 1.$$

Сравнение с (8) показывает, что универсальное хеширование по базису двух рациональных функций в два раза выигрывает по сложности вычислений.

3. Зафиксируем вероятность коллизии значением $\varepsilon = \rho_{g+1} / N = 2g / q^2$. Для условия $\rho_k = \rho_{g+1}$ Число слов данных при хешировании по полному функциональному базису определяется родом кривой

$$k_{xyvw} = qq_0 - q_0 + 1. \quad (14)$$

При хешировании по двух параметрическому базису функционального поля кривой Судзуки наибольшее число слов данных $\rho_k \le \rho_{g+1}$ определяется соотношением (5)

$$k_{xy} = (2q_0 - 1)(q_0 + 1). \quad (15)$$

Отношение числа хешируемых слов данных при двух параметрическом хешировании к хешированию по полному базису равно

$$R = k_{xy} / k_{xyvw} = (2q_0 - 1)(q_0 + 1)/(qq_0 - q_0 + 1) \approx 1/q_0 \quad (16)$$

и определяет проигрыш двух параметрического хеширования при фиксированной вероятности коллизии.

Выводы

1. Универсальное хеширование с ограничением функционального поля алгебраической кривой приводит к существенному снижению сложности вычислений. При этом уменьшается число хешируемых данных. Требуется оптимизация базиса функционального поля. Эффект достигается на сложных многопараметрических кривых.

2. Уменьшение размерности функционального поля алгебраической кривой при рациональном выборе базисных функций приводит к схемам хеширования, которые вкладываются в хеширование по наилучшим плоским кривым. Так двух параметрическое хеширование по кривой Судзуки по алгоритму вычисления, асимптотическим оценкам хеширования приводит к определению хеширования по кривой Эрмита. Можно показать, что хеширование по одной базисной функции приводит к хеш функции по проективной прямой.

3. Предложен метод универсального хеширования с ограничением функционального поля алгебраической кривой, который является дальнейшим развитием метода универсального хеширования на основе скалярного произведения по рациональным функциям линейного базисного пространства и отличается от известного вычислением хеш функций по подмножеству рациональных функций функционального поля с упорядоченными порядками полюсов.

Список литературы: 1. *Халимов, Г.З.* Аутентификация с применением Эрмитовых кодов / Г.З. Халимов, А.Ю. Иохов // Вестник ХПИ. – Х., 2005. – Вып. 9. – С. 26-32. 2. *Халимов, Г.З.* Универсальное хеширование по максимальным кривым Гурвица / Г.З. Халимов // Прикладная радиоэлектроника. – 2010. – Т.9, № 3. – С.365-370. 3. *Халимов, Г.З.* Универсальное хеширование по кривым Сузуки / Г.З.Халимов, Е.В.Котух // Прикладная радиоэлектроника. –2011. – Т. 10, № 2. –С.164-170. 4. *Deligne, P.* Representations of reductive groups over finite fields / P.Deligne, Lusztig // Annatatsiya of Mathematics. – 1976. – N.103. – P.103–161. 5. *Халимов, Г.З.* Алгоритм универсального хеширования по кривой Сузуки / Г.З.Халимов, Е.В.Котух // Восточно-Европейский журнал передовых технологий. – 2011. – № 3/9 (51). – С.10-16. 6. *Халимов, Г.З.* Оценка параметров кривых Ферма для универсального хеширования в простом поле / Г.З.Халимов // Науч.-техн. конф. с междунар. участием. Компьютерное моделирование в наукоемких технологиях (часть 2). КМНТ Харьков, 18-21 мая 2010. – С.266. 7. *Fuhrmann, R.* The genus of curves over finite fields with many rational points / R.Fuhrmann, F.Torres // Manuscripta Mathematics. – 1996. – N.89. – P.103–106. 8. *Torres, F.* The Deligne-Lusztig curve associated to the Suzuki group [Электронный ресурс]/ F.Torres // arXiv:alg-geom/9706012v1 26 June 1997. 9. *Халимов, Г.З.* Универсальное хеширование по рациональным функциям кривой Эрмита / Г.З.Халимов, А.Ю.Иохов // Междунар. науч.-практ. конф. «Застосування інформаційних технологій у підготовці та діяльності сил охорони правопорядку». Академія внутрішніх військ МВС України 17-18.03.2011. Зб. тез доповідей. – 2011. – С.48-51.

Харьковский национальный университет радиоэлектроники

Поступила в редколлегию 15.09.2012

УДК 004.056.55

І.Д. ГОРБЕНКО, д-р техн. наук, М.В. ЄСІНА

АЛГОРИТМ РЕШЕТА ЧИСЛОВОГО ПОЛЯ

Вступ

В той час, коли був винайдений RSA, числа більш ніж з 80 десятковими знаками не піддавалися розкладанню. Всі відомі алгоритми або працювали дуже повільно, або вимагали чисел спеціального вигляду. Це робило відносно безпечними навіть маленькі, 256-бітові ключі.

Першим серйозним проривом в процедурі розкладання числа (факторизації) було квадратичне решето, quadratic sieve (QS). Це відносно простий алгоритм факторизації, запропонований Carl Pomerance в 1981 р., який може розкладати на множники числа до 110 десяткових розрядів чи приблизно таких і для таких чисел залишається кращим. Для ще більших чисел застосовується метод решета числового поля, general number field sieve (GNFS). Проте метод решета числового поля навіть для базового опису вимагає складних багатобічних роз'яснень і обґрунтувань. В той же час основні ідеї обох методів решета збігаються.

Поява на початку 90-х років алгоритму решета числового поля (NFS) для вирішення завдання факторизації і доказ для нього асимптотично найкращої оцінки трудомісткості практично припинила подальші дослідження в області алгоритмів квадратичного решета [5].

Алгоритм решета числового поля для факторизації цілих чисел спеціального вигляду (special number field sieve, або SNFS) був вперше запропонований в 1990 р. За його допомогою було розкладено на множники число Ферма $F_9 = 2^{512} + 1$, що записувалось 155 десятковими знаками. Евристична оцінка складності складає $L_N[1/3;\ c]$ арифметичних операцій при $c = (32/9)^{1/3} = 1{,}5263...$ [1].

Згодом метод був узагальнений і застосований для факторизації довільних цілих чисел. Він отримав назву general number field sieve або, скорочено, GNFS. Оцінка складності також складає $L_N[1/3;\ c]$ при деякій постійній c.

На сьогоднішній день алгоритми решета числового поля і квадратичне решето Померанца – найшвидші з відомих алгоритмів факторизації великих чисел. Алгоритм решета числового поля, запропонований Широкауером, при $p > 10^{100}$ працює ефективніше за різні модифікації методу COS; його часова складність складає порядку $L_N\left[1/3;\ (64/9)^{1/3}\right]$ арифметичних операцій [1].

В даний час рекордним значенням для натуральних чисел, розкладених за допомогою SNFS, є число спеціального вигляду, що записане 227 десятковими знаками. Для RSA-чисел n, що не мають спеціального вигляду, рекордні розкладання були знайдені у 1999 р. спочатку було розкладено 140-значне RSA-число, а потім – 155-значне RSA-число (512 бітів). На факторизацію цього останнього числа було потрібно близько 8400 mips-year [1].

Оцінка криптостійкості RSA-129 авторами алгоритму складає приблизно 40 квадриллі онов років роботи ЕОМ (4 1016 MIPS-років).

У таблиці наведено результати факторизації чисел RSA – k [2].

QS (Quadratic Sieve) – метод квадратичного решета;

GNFS (General Number Field Sieve) – узагальнений метод решета числового поля.

Число	Дата	Складність, MIPS-років	Алгоритм
RSA – 100	Квітень 1991	7	QS
RSA – 110	Квітень 1992	75	QS
RSA – 120	Червень 1993	830	QS
RSA – 129	Квітень 1994	5000	QS
RSA – 130	Квітень 1996	500	GNFS
RSA – 140	Лютий 1999	2000	GNFS
RSA – 155	Серпень 1999	8000	GNFS

Мета статті – розглянути реалізацію методу загального решета числового поля та особ-вості факторизації на основі даного методу. Визначити вибір побудови факторних баз та значити правила побудови алгебраїчної факторної бази. Розглянути побудову алгебраїчної зи характерів та оцінити складність решета числового поля. Задача статті – обґрунтувати бір факторних баз. Визначити принцип побудови алгебраїчної факторної бази та алгебраї-ої бази характерів. Оцінити складність решета числового поля.

Особливості факторизації на основі загального «решета числового поля»

Етапи факторизації на основі використання загального решета числового поля:
- вибирання поліномів відповідних степенів;
- просіювання з відбиранням позитивних даних;
- обробка даних з розв'язанням задачі лінійної алгебри;
- знаходження нетривіальних рішень.

Метод загального решета числового поля (NSF – number field sieve) дозволяє фактори-увати модуль RSA перетворення зі складністю (асимптотичною):

$$L_N(\gamma,\ \delta) = \exp(\delta(Ln(N))^{\gamma}\, Ln\big(Ln(N)\big)^{(1-\gamma)}, \qquad (1)$$

де $\gamma = 1/3$, а $\delta = (64/9)^{1/3} T$ (приблизно 1.923) параметри методу.

Для методу спеціального решета числового поля (SNSF – special number field sieve) параметри методу дорівнюють

$$\gamma = 1/3, \text{ а } \delta = (32/9)^{1/3} \text{ (приблизно 1,526)}, \qquad (2)$$

тобто метод спеціального решета числового поля є менш складним (більш швидкодіючим).

Взагалі ідея методу загального решета числового поля належить Джону Полларду, який 1988 р. запропонував просіювання виконувати не у кільці цілих чисел, як це робиться в квадратичному решеті, а в алгебраїчному полі. Спочатку метод можна було використовувати для факторизації тільки чисел спеціального вигляду $2^n \pm n$. Тому метод отримав назву «спеціального решета числового поля». Практична реалізація ідеї Полларда була здійснена в 990 р., коли з його використанням було факторизовано число Ферма (2^{512}). Також були факторизовані деякі числа вигляду $b^c \pm 1$. У подальшому було запропоновано використовувати метод решета числового поля й для факторизації довільних цілих чисел. Була знайдена евристична оцінка його складності. Тобто множник δ був зменшений у порівнянні з квадратичним решетом з 1/2 до 1/3 [3].

Розглянемо етапи базового методу решета числового поля [4].

Нехай n – непарне ціле число, яке необхідно факторизувати. Основна ідея Полларда полягає в тому, щоб замінити поліном 2-го степеня $q(x) = (x+m)^2 - n$, який використовувався у квадратичному решеті, на довільний поліном $P_d(x)$ степеня $d \geq 3$, який задовольняє умові $P_d(m) = n$ для деякого цілого числа m. Далі, просіювання за множиною цілих чисел Z було замінене просіюванням в кільці $Z(\beta)$, яке отримується приєднанням до кільця Z цілого

алгебраїчного числа β, що є коренем полінома $P_d(m)$. У порівнянні з квадратичним решетом, у решеті числового поля факторна база складається із простих елементів кільця алгебраїчних чисел.

Виграш при використанні решета числового поля полягає в тому, що умова $P_d(m) = n$ для деякого цілого m, що накладається на поліном $P_d(x)$, у порівнянні з коефіцієнтами, що використовуються у квадратичному решеті, може бути виконана при менших значеннях коефіцієнтів полінома $P_d(x)$.

Реалізація методу загального решета числового поля [4]:

1) Обирається степінь незвідного полінома $d \geq 3$. Можна взяти $d = 2$, але в цьому випадку у порівнянні з квадратичним решетом виграшу не буде.

2) Обирається ціле число m – таке, що $\left\lfloor n^{1/(d+1)} \right\rfloor < m < \left\lfloor n^{1/d} \right\rfloor$ та розкладається число n за основою m, тобто подається у вигляді

$$n = m^d + a_{d-1}m^{d-1} + \ldots + a_0 \ \ldots \tag{3}$$

3) Із розкладом (3) пов'язується незвідний поліном у кільці $Z(x)$:

$$f_1(x) = x^d + a_{d-1}x^{d-1} + \ldots + a_0 \ldots \tag{4}$$

4) Визначається поліном просіювання $F_1(a, b)$ як однорідний поліном від двох змінних a та b:

$$F_1(a, b) = b^d f_1(a/b) = a^d + a_{d-1}a^{d-1}b + a_{d-2}a^{d-2}b^2 + \ldots + a_0 b^d \tag{5}$$

Необхідно відмітити, що значення $F_1(a,b)$ дорівнює нормі полінома $a - b \times x$ в алгебраїчному числовому полі $Q[\beta]$, яке отримують доповненням поля раціональних чисел Q у загальному випадку комплексного кореня β багаточлена $f_1(x)$. При цьому властивості комутативності норми

$$Nr\left(h_1(x) \times h_2(x)\right) = \left(Nr\left(h_1(x)\right) \times \left(Nr\left(h_2(x)\right)\right)\right) \tag{6}$$

дозволяють замість розкладання багаточлена з кільця $Z(\beta)$ виконати розкладання їх норм.

5) Визначається другий поліном

$$f_2(x) = x - m \tag{7}$$

та відповідний йому однорідний поліном

$$F_2(a, b) = a - bm. \tag{8}$$

Головною вимогою при виборі пари поліномів $(f_1(x), f_2(x))$ є виконання вимоги

$$f_1(m) = f_2(m) (mod\ n), \tag{9}$$

яка в нашому випадку, очевидно, виконується, оскільки перший багаточлен у точці m дорівнює n, а другий – нулю.

6) Обирається два позитивних числа L_1 та L_2, що визначають деяку прямокутну область

$$SR = \{1 \leq b \leq L_1, -L_2 \leq a \leq L_2\}, \tag{10}$$

яку називають областю просіювання.

7) Нехай β – корінь полінома $f_1(x)$. Розглядається кільце багаточленів $Z(\beta)$ (для формального описання алгоритму). Також визначимо алгебраїчну факторну базу FB_1, що скла-

дається з поліномів першого порядку вигляду $a-b\times\beta$ з нормою, що є простим числом. Такі багаточлени є простими елементами, що не розкладаються, в кільці алгебраїчних цілих поля $K=Q(\beta)$. Абсолютні величини норм багаточленів із факторної бази FB_1 обмежимо зверху деякою постійною B_1.

8) Визначається раціональна факторна база FB_2, яка складається з усіх простих чисел, добуток яких обмежується другою постійною B_2.

9) Визначається невелика множина поліномів першого порядку $c-d\times\beta$, норма яких є простим числом. Позначимо цю множину як FB_3. Вона має задовольняти умові, що $FB_1 \wedge FB_3=\varnothing$ і називається факторною базою квадратичних характерів. Факторна база FB_3 необхідна на підсумковій стадії алгоритму для перевірки того факту, що знайдений у ході просіювання багаточлен є повним квадратом.

10) Далі для отримання гладких пар (a,b) множини M здійснюється просіювання поліномів $\{a-b\times\beta\,|\,(a,b)\in SR\}$ згідно з факторною базою FB_1, а також цілих чисел $\{a-b\times m\,|\,(a,b)\in SR\}$ згідно з факторною базою FB_2. При цьому пара (a,b) називається гладкою, якщо $НСД(a,b)=1$, а поліном $a-b\times\beta$ і число $a-b\times m$ розкладається по відповідним факторним базам FB_1 та FB_2. При цьому число гладких пар у множині M має бути більше загальної суми елементів усіх трьох баз щонайменше на дві одиниці.

11) Далі шукають підмножину $S\supset M$ таку, що добуток усіх пар $\prod_{(a,b)\in S}Nr(a-b\times\beta)=\mathrm{H}^2$ для $\mathrm{H}\in Z$, а також $\prod_{(a,b)\in S}Nr(a-b\times m)=B^2$, $B\in Z$. Для знаходження множини S, як і в методі квадратичного решета, складається система лінійних алгебраїчних рівнянь з коефіцієнтами із множини $F_2=\{0,1\}$, результатом розв'язання якої й будуть номери S.

12) Формується поліном

$$g(\beta)=\left(f_1'(\beta)\right)^2\prod_{(a,b)\in S}(a-b\times\beta) \tag{11}$$

де $f_1'(x)$ – похідна полінома $f_1(x)$.

13) Якщо вся процедура виконана коректно, то поліном $g(\beta)$ є повним квадратом у кільці поліномів $Z(\beta)$. Знаходимо квадратні корені із полінома $g(\beta)$ та цілого числа B^2, внаслідок чого знаходимо поліном $\alpha(\beta)$ та число B.

14) Замінимо поліном $\alpha(\beta)$ на число $\alpha(m)$. Відображення $\varphi\colon \beta\to m$ є кільцевим гомоморфізмом кільця алгебраїчних цілих чисел Z_κ у кільце Z. Звідки отримаємо співвідношення:

$$\begin{aligned}A^2=g(m)^2=(\varphi(g(x)^2))=\varphi\left(f_1'(\beta)\right)^2\prod_{(a,b)\in S}(a-b\times\beta)=\\ =\left(f_1'(m)\right)^2\prod_{(a,b)\in S}(a-b\times m)=\left(f_1'(m)\right)^2C^2(mod n)\end{aligned} \tag{12}$$

Таким чином, визначивши $B=f'(m)C$, знайдемо пару цілих чисел (A,B), які задовольняють умові

$$A^2=B^2(mod\ n) \tag{13}$$

На останок можна знайти дільник числа n, обчислюючи $НСД(n,A\pm B)$.

Вибір факторних баз

Розглянемо спочатку процедуру побудови раціональної факторної бази FB_2. Ця база використовуватиметься для розкладання чисел вигляду $a-bm$ в множині Z, тому множина FB_2 визначається рівною множині всіх простих чисел, обмежених зверху константою B_1. Границя B_1 для рекордних розкладань доходить до гігантських значень 10^6-10^7 [4].

Побудова алгебраїчної факторної бази

Алгебраїчна факторна база складається з лінійних многочленів $c-d\theta$, що породжують прості ідеали в кільці цілих чисел алгебри Z_K. Побудова такої факторної бази є дуже складним завданням, проте, наступна теорема дозволяє перейти від многочленів, що не наводяться, і породжуваних ними простих ідеалів до пар натуральних чисел $(p,\ r)$ [4]:

Теорема 1. Безліч простих ідеалів кільця Z_K знаходиться у взаємно-однозначній відповідності з безліччю пар позитивних цілих чисел $(p,\ r)$ таких, що p – просте число, $0\le r<p$, та $f_1(r)\equiv 0(\mathrm{mod}p)$. Завдяки цьому результату, можна не виписувати явно прості ідеали кільця Z_K, а просто встановити границю B_1 і шукати всі пари $(p,\ r)$, де $p\le B_1$ – просте число, а $r\in[0,p-1]$ та $f_1(r)\bmod \mathrm{p}=1$.

Простий перебір всіх можливих пар працює неефективно для великих чисел p. Опишемо один алгоритм, що дозволяє поліпшити пошук коріння многочлена $f_1(x)\bmod p$. Цей алгоритм ґрунтується на наступній теоремі:

Теорема 2. У кінцевому полі $GF_q=GF_{p^k}$ поліном x^q-x повністю розкладається на лінійні множники

$$x^q-x=\prod_{i=0}^{q-1}(x-i). \tag{14}$$

Алгоритм обчислення коріння за модулем простого числа p працює таким чином [4]:

1. Шукаємо $g(x)=H.O.Д.(f_1(x),x^{\mathrm{p}}-x)$. Метою цього кроку є відсікання частини полінома $f_1(x)$, що має корінь за модулем p. Наприклад, якщо виявиться, що $g(x)=1$, тоді $f_1(x)$ не має коренів за модулем p.

2. За теоремою 2 для будь-якого b, $0\le b<p$

$$g(x-b)\,|\,x^p-x \text{ та } x^p-x=x(x^{(p-1)/2}+1)(x^{(p-1)/2}-1) \tag{15}$$

3. Використовуючи властивість 15, відокремимо корені $g(x)\bmod p$.

Розглянемо цей алгоритм на прикладі полінома $f_1(x)=x^3+15x^2+29x+8$, $p=67$:

1) Обчислимо $g(x)=H.O.Д.\left(f_1(x),x^{\mathrm{p}}-x\right)=H.O.Д.(f_1(x),x^{67}-x=x^3+15x^2+29x+8)$. Отримаємо $g(x)=f_1(x)$, отже, $f_1(x)$ містить всі три можливі корені $f_1(x)$ за модулем 67.

2) При $b=0$ рівність (15) отримає вигляд $g(x)=x^3+15x^2+29x+8\,|\,x\left(x^{33}+1\right)\left(x^{33}-1\right)$.

Оскільки $g(0)=8\ne 0$, $x=0$ не є коренем $g(x)$.

3) Обчислимо $H.O.Д.(\mathrm{g}(x),x^{33}+x)=x^2+21x+21$, та $H.O.Д.(g(x),x^{33}-x)=x+61$, звідки знайдемо корінь $x=-61\equiv 6(\mathrm{mod}\ 67)$. Два корені, що залишилися, є коренями полінома $g_1(x)=x^2+21x+21\equiv x^2+21x-46(\mathrm{mod}67)$. Їх можна отримати, наприклад, обчислюючи за допомогою алгоритму Шенкса -- Тонеллі квадратний корінь з дискримінанта цього рівняння $D=(-21)^2+4\cdot 46=2998\equiv 50(\mathrm{mod}67)$.

Повторимо це обчислення з іншими значеннями b. Підставляючи в рівність (15) $b=1$, отримаємо $Н.О.Д.(g(x-1), x^{33}+x) = g(x-1)$ та $Н.О.Д.(g(x-1), x^{33}-x) = 1$. Значить, розщеплення $g_1(x_1)$ не відбувається і $b=1$ не вирішує нашої задачі.

Підставимо $b=2$. Отримаємо $Н.О.Д.(g(x-2), x^{33}+x) = x+21$ та $Н.О.Д.(g(x-2), x^{33}-x) = x+63$, звідки $x=-21\equiv 46(\mathrm{mod}\ 67)$, $x=-63\equiv 6(\mathrm{mod}\ 67)$. Числа $x=44$ і $x=2$ є коренями $f_1(x)(\mathrm{mod}\ 67)$. Всі три корені знайдено.

Побудова алгебраїчної бази характерів

Побудова третьої факторної бази, що складається з квадратичних характерів, виконується також, як і алгебраїчної. Вибирається новий обмежувач $B_3 > B_2$ і розглядаються всі прості числа від B_2 до B_3. Для рекордних розкладань B_3 вибирається так, щоб в інтервал $[B_2, B_3]$ потрапило $10^4 - 10^5$ простих чисел. Вибір розмірності B_3 визначає міру впевненості в тому, що знайдений в результаті просіювання поліном є повним квадратом [4].

Оцінка складності решета числового поля

Загальний час роботи алгоритму решета числового поля залежить від часу роботи складових його частин, з яких найбільшу вагу мають час первинного просіювання, в ході якого шукаються номери гладких пар для складання системи лінійних рівнянь і час обчислення квадратного кореня з полінома в просторі $Z[x]/(f_1(x))$. Всі останні складові алгоритму впливають значно менше на продуктивність GNFS [3].

Наведемо оцінку методу решета числового поля, узяту на основі функції $L_N(\alpha;c)$: $L_N(\alpha;c) = \exp(c+o(1))(\ln n)^{\alpha}(\ln\ln n)^{1-\alpha}$. Ця оцінка виконана за умови, що степінь d і границя області просіювання у обрані, як вказано нижче [4]:

$$d = 3^{1/3} + o(1)\left(\log n / \log\log n\right)^{1/3},\ n > d^{2d^2} > 1,\ u = y = L_N(1/3,\ (8/9)^{1/3} + o(1)). \qquad (16)$$

Якщо операції множення поліномів і обчислення лишків за модулем $f(x)$ виконані за допомогою дискретного перетворення Фур'є, то час обчислення квадратного кореня визначається за допомогою наступної оцінки:

$$T(n) = y_1 + o(1), \qquad (17)$$

де y – верхня границя для параметрів a, b області просіювання, залежна від числа n і степені d полінома $f_1(x)$. Її оптимальне значення

$$\log y = (1/2 + o(1))(d\log s + \sqrt{(d\log d)^2 + 4\log(n^{1/d})\log\log(n^{1/d})}) \qquad (18)$$

При виконанні умов (16) і часу обчислення кореня (18) час роботи алгоритму решета числового поля оцінюється величиною

$$T(n) = L_N\left(1/3,\ (64/9)^{1/3} + o(1)\right). \qquad (19)$$

Відзначимо, що наближене значення константи $(64/9)^{1/3}$ дорівнює 1,92. Таким чином, зменшення значення показника степеня в найбільш важливому співмножнику $\log n$ функції $L_N(\alpha;c)$ від значення 1/2 в методі квадратичного решета до 1/3 в методі решета числового поля дає той прогрес, який забезпечує пріоритет цього методу над методом квадратичного решета і всіма іншими методами факторизації, відомими на сьогоднішній день [3].

Висновки

Якщо ми хочемо факторизувати натуральне число n, то спочатку перебором $p = 2,3,5,7...$ до деякої границі слід відокремити маленькі прості дільники нашого числа. Потім слід перевірити, чи є наше число, яке ми хочемо факторизувати, складеним. Для цього краще всього використовувати імовірнісний тест Міллера – Рабіна. Якщо наше число – ймовірно просте, то потрібно спробувати довести його простоту за допомогою алгоритму Ленстри – Коєна. Якщо наше число – складене, то можна спробувати отримати його розкладання на множники за допомогою (P-1)-метода Полларда і ρ-методу Полларда, а також за допомогою методу еліптичних кривих Ленстри. Після цього для факторизації слід застосувати метод квадратичного решета, якщо наше число n не перебільшує 10^{110}. Для чисел більшої величини слід використовувати алгоритми решета числового поля. Стосовно криптографії з відкритим ключем бачимо, що RSA-модулі n, що дорівнюють добутку двох простих чисел, не є безпечними для шифрування за умови $n \approx 2^{512}$.

RSA-модулі $n \approx 2^{1024}$ залишатимуться безпечними ще принаймні 15 років, якщо не будуть знайдені принципово нові алгоритми факторизації або не буде створений ефективний квантовий комп'ютер.

На даний час, метод решета числового поля – найефективніший метод факторизації для чисел $n > 10^{110}$. Фактично решето числового поля не є алгоритмом. Це метод обчислення, що складається з декількох етапів, і кожен з цих етапів обслуговується декількома алгоритмами.

Список литературы: 1. *Василенко, О.Н.* Теоретико-числовые алгоритмы в криптографии. Москва: МЦНМО, 2003. 328 с. 60. 2. *Воронков, Б.Н.* Криптографические методы защиты информации : учеб. пособие для вузов. – Воронеж : Изд.-полигр.центр Воронеж. гос. ун-та, 2008. – 60 с. 15. 3. *Горбенко, І.Д., Горбенко, Ю.І.* Прикладна криптологія. – Харків : Форт, 2012. – 867. 4. *Ишмухаметов, Ш.Т.* Методы факторизации натуральных чисел. – Казань : Казанск. ун-т, 2011. – 202. 5. *Математика* и безопасность информационных технологий // Материалы конференции в МГУ 28 – 29 октября 2004 г. – М. : Изд-во МЦНМО, 2005. – С. 192-199.

Харківський національний
університет радіоелектроніки

Надійшла до редколегії 19.09.2012

УДК 004.056.55

В. Ю. КОВТУН, канд. техн. наук, А. А. ОХРИМЕНКО

ПОДХОДЫ К РАСПАРАЛЛЕЛИВАНИЮ ПРОГРАММНОЙ РЕАЛИЗАЦИИ ОПЕРАЦИИ УМНОЖЕНИЯ ЦЕЛЫХ ЧИСЕЛ

Информатизация общества ведет к повышению роли систем защиты информации, которые немыслимы без криптографических преобразований, среди последних особое место занимают криптографические преобразования с открытым ключом.

Криптографические преобразования с открытым ключом имеют продолжительную историю: от публикации Диффи и Хеллмана [1], которая заложила фундамент, до современных криптосистем, например на алгебраических кривых. Среди актуальных задач дальнейшего развития криптосистем с открытым ключом выделяют повышение производительности программной и аппаратной реализации. На протяжении всего периода развития криптопреобразований с открытым ключом их основу составляли операции в кольцах и полях чисел, среди которых умножение занимает особое место.

Цель данной работы – поиск путей увеличения скорости преобразований с открытым ключом путём увеличения быстродействия операций умножения в кольцах и полях чисел методом отложенного переноса и эффективного распараллеливания.

Программная реализация любого алгоритма напрямую зависит от архитектуры аппаратной платформы. Развитие микропроцессорной техники до некоторого времени шло в сторону увеличения тактовой частоты процессоров, однако после достижения физического предела частоты акцент был сделан на увеличение числа потоков обработки команд. На сегодняшний день процессоры имеют от 2 до 10 физических ядер, но эффективность использования всех вычислительных возможностей процессоров всецело ложиться на разработчика, который проектирует и программирует алгоритмы.

В связи с этим актуальность адаптации существующих алгоритмов на процессорах с несколькими потоками выполнения команд не вызывает сомнений.

Задача распараллеливания алгоритмов арифметических операций известна давно [2, 3]. В этих работах рассматриваются алгоритмы умножения Монтгомери и целочисленной арифметики для реализации на NVIDIA GPGPU [2, 3]. Дальнейшее развитие данного направления для других алгоритмов умножения позволит найти наиболее эффективные техники распараллеливания для различных аппаратных платформ.

Известны следующие технологии распараллеливания:

- OpenMP [4, 5] – для процессоров общего назначения;
- OpenCL [6] – для процессоров общего назначения и графических процессорах NVIDIA и AMD;
- Intel Threading Building Block [7] – для процессоров общего назначения;
- NVIDIA CUDA [8] – для графических процессорах общего назначения NVIDIA;
- AMD Accelerated Parallel Processing (APP) [9] графических процессорах общего назначения AMD.

Далее рассмотрим алгоритмы умножения целых чисел, а также подходы к их распараллеливанию с помощью технологии OpenMP. Технология OpenMP была выбрана не случайно - она поддерживается большинством современных компиляторов языка C++ для различных аппаратных платформ, а также в связи с ее доступностью, простотой и наглядностью. Другие технологии более громоздки и менее наглядны, при этом суть предложенного подхода к распараллеливанию остается неизменной.

Описание алгоритма Modified Comba

Ранее, в работе [11], авторами был предложен модифицированный алгоритм Comba [10] - Modified Comba, в котором используется подход отложенного переноса. Использование

64-разрядных переменных для хранения 32-разрядных переменных позволило избавиться от учета переноса из 32-разрядной переменной после каждой арифметической операции.

Перенос накапливался в старшей части 64-разрядной переменной и может быть учтен при необходимости (рис. 1). Алгоритм Modified Comba [11] приведен ниже.

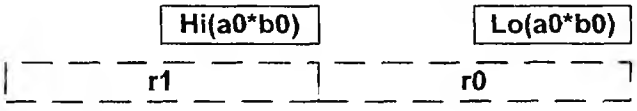


Рис. 1. Идея механизма отложенного переноса

Input: *integers* $a, b \in \mathrm{GF}(p)$, $w = 32$, $n = \log_{2^w} a$, $nk = 2n - 1$.

Output: *integer* $c = a \cdot b$.

1. $r_0^{(64)} \leftarrow 0$, $r_1^{(64)} \leftarrow 0$, $r_2^{(64)} \leftarrow 0$.
2. For $k \leftarrow 0$, $k < n$, $k{+}{+}$ do
2.1. For $i \leftarrow 0$, $j \leftarrow k$, $i \le k$, $i{+}{+}$, $j{-}{-}$ do
2.1.1. $(uv)^{(64)} \leftarrow a_i^{(32)} \cdot b_j^{(32)}$.
2.1.2. $r_0^{(64)} \leftarrow r_0^{(64)} + v^{(32)}$, $r_1^{(64)} \leftarrow r_1^{(64)} + u^{(32)}$.
2.2. $r_1^{(64)} \leftarrow r_1^{(64)} + \mathrm{hi}_{(32)}\left(r_0^{(64)}\right)$, $r_2^{(64)} \leftarrow r_2^{(64)} + \mathrm{hi}_{(32)}\left(r_1^{(64)}\right)$.
2.3. $c_k^{(32)} \leftarrow \mathrm{low}_{(32)}\left(r_0^{(64)}\right)$, $r_0^{(64)} \leftarrow \mathrm{low}_{(32)}\left(r_1^{(64)}\right)$, $r_1^{(64)} \leftarrow \mathrm{low}_{(32)}\left(r_2^{(64)}\right)$, $r_2^{(64)} \leftarrow 0$.
3. For $k \leftarrow n$, $l \leftarrow 1, k < nk$, $k{+}{+}$, $l{+}{+}$ do
3.1. For $i \leftarrow l$, $j \leftarrow k - l$, $i < n$, $i{+}{+}$, $j{-}{-}$ do
3.1.1. $(uv)^{(64)} \leftarrow a_i^{(32)} \cdot b_j^{(32)}$.
3.1.2. $r_0^{(64)} \leftarrow r_0^{(64)} + v^{(32)}$, $r_1^{(64)} \leftarrow r_1^{(64)} + u^{(32)}$.
3.2. $r_1^{(64)} \leftarrow r_1^{(64)} + \mathrm{hi}_{(32)}\left(r_0^{(64)}\right)$, $r_2^{(64)} \leftarrow r_2^{(64)} + \mathrm{hi}_{(32)}\left(r_1^{(64)}\right)$.
3.3. $c_k^{(32)} \leftarrow \mathrm{low}_{(32)}\left(r_0^{(64)}\right)$, $r_0^{(64)} \leftarrow \mathrm{low}_{(32)}\left(r_1^{(64)}\right)$, $r_1^{(64)} \leftarrow \mathrm{low}_{(32)}\left(r_2^{(64)}\right)$, $r_2^{(64)} \leftarrow 0$.
4. $c_{nk}^{(32)} \leftarrow \mathrm{low}_{(32)}\left(r_0^{(64)}\right)$.
5. Return (c).

Проанализируем алгоритм Modified Comba [11], укажем его основные отличия от прототипа – алгоритма Comba [10], более детально остановимся на его потенциальных возможностях.

На рис. 2 и 3 приведена графическая интерпретация алгоритма Modified Comba для $n = 3$, где четко прослеживается сложение результатов соответствующих произведений по столбцам.

Изложенная в алгоритме Modified Comba [11] идея отложенного переноса натолкнула авторов на возможность параллельно выполнить сложение значений в столбцах $r_0 = \sum_{k=0}^{2n-1} Lo\left(a_i \cdot b_j\right) \mid k = i + j, 0 \le i, j < n$ и $r_1 = \sum_{k=0}^{2n-1} Hi\left(a_i \cdot b_j\right) \mid k = i + j, 0 \le i, j < n$.

В классическом алгоритме Comba этот подход неосуществим из-за связанности операций сложения переносом из старшего разряда. Факт отсутствия переноса при сложении чисел в столбце (накоплении суммы) для Modified Comba позволяет говорить об изолированности операции накопления суммы, что в свою очередь дает возможность выполнять цикл накопления п.2 и 3 параллельно, в отдельных (независимых) потоках.

Заметим, что после завершения накопления суммы в отдельном потоке все же необходимо выполнить корректировку (учесть перенос) $r_1 = r_1 + Hi(r_0)$, $r_2 = r_2 + Hi(r_1)$ и сформировать результат $c_i = Lo(r_0)$.

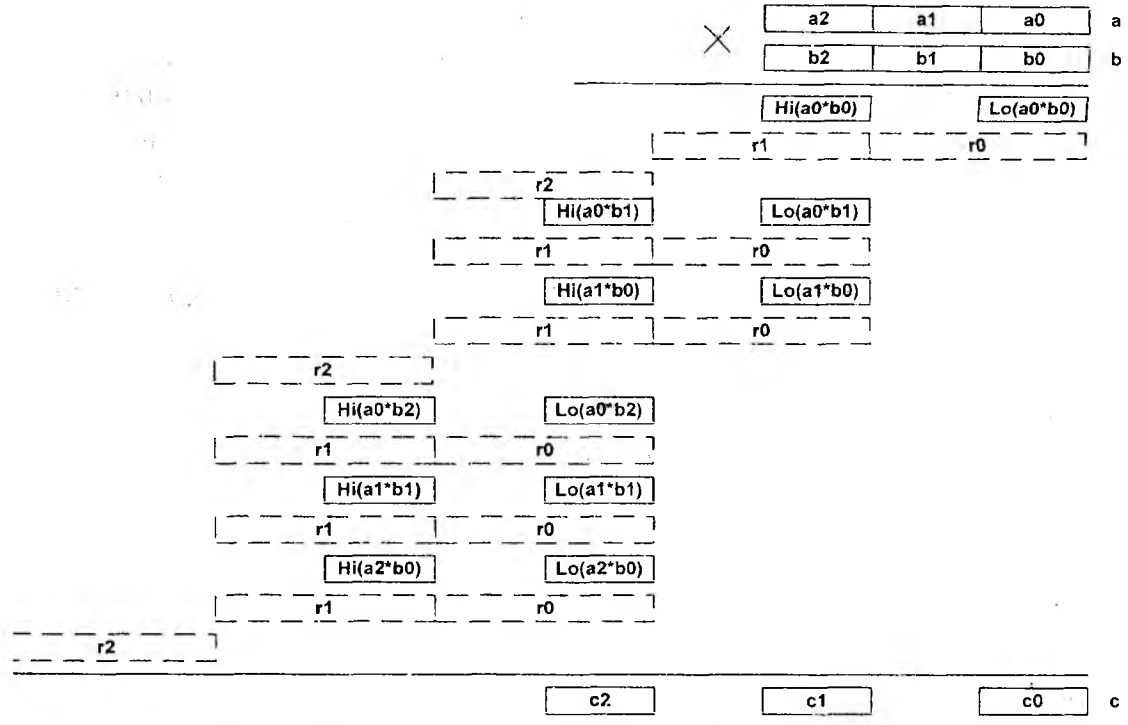


Рис. 2. Графическая интерпретация цикла 2, алгоритма Modified Comba

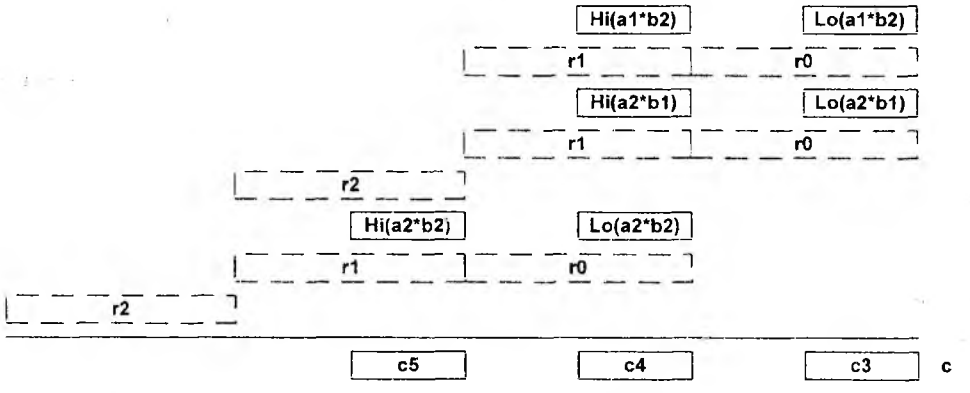


Рис. 3. Графическая интерпретация цикла 3, алгоритма Modified Comba

Описанный механизм отложенного переноса позволяет сформулировать несколько подходов к распараллеливанию алгоритма умножения Modified Comba:

- Параллельное выполнение циклов п. 2 и п. 3, с последующей коррекцией результатов (два параллельных потока). В дальнейшем будем называть его Modified Comba 2x;
- Параллельное выполнение итераций циклов п. 2 и п. 3, с последующим слиянием промежуточных результатов (множество параллельных потоков). В дальнейшем будем называть его Modified Comba Mx.

Описание алгоритма Modified Comba 2x с двумя потоками

В алгоритме присутствует два цикла п. 2 и 3, которые производят чтение элементов $a_i^{(32)}$ и $b_j^{(32)}$ соответствующих массивов, а также запись результатов умножения в элементы $c_k^{(32)}$. Заметим, что индексы k в циклах п. 2 и 3 не повторяются при записи в $c_k^{(32)}$, это позволяет говорить о независимости данных в циклах и о возможности применять распараллеливание для их параллельного выполнения. Следует обратить внимание, что оба цикла п. 2 и 3 используют общие временные переменные r_0, r_1 и r_2. Причем, в r_0 и r_1 хранятся

значения, которые используются в цикле п. 3 после окончания цикла п. 2. Таким образом, после окончания цикла п. 3 необходимо выполнить корректировку – учесть результаты работы цикла п. 2 в результатах цикла п. 3, которые хранятся во временных переменных r_0 и r_1. Заметим, что при распараллеливании каждый поток работает с собственными временными переменными rl_0, rl_1 и rl_2. Глобальные переменные r_0 и r_1 необходимы лишь для передачи возможного переноса из цикла п. 2 для последующей корректировки результатов накопления в цикле п. 3.

Рассмотрим алгоритм Modified Comba с распараллеливанием с помощью OpenMP на два потока:

Input: integers $a, b \in \mathrm{GF}(p)$, $w = 32$, $n = \log_{2^w} a$, $nk = 2n-1$.

Output: integer $c = a \cdot b$.

1. #pragma omp parallel sections private($r_0^{(64)}, r_1^{(64)}$) begin

1.1. #pragma omp section begin

1.1.1. $rl_0^{(64)} \leftarrow 0$, $rl_1^{(64)} \leftarrow 0$, $rl_2^{(64)} \leftarrow 0$.

1.1.2. For $k \leftarrow 0$, $k < n$, $k{+}{+}$ do

1.1.2.1. For $i \leftarrow 0$, $j \leftarrow k$, $i \le k$, $i{+}{+}$, $j{-}{-}$ do

1.1.2.1.1. $(uv)^{(64)} \leftarrow a_i^{(32)} \cdot b_j^{32}$.

1.1.2.1.2. $rl_0^{(64)} \leftarrow rl_0^{(64)} + v^{(32)}$, $rl_1^{(64)} \leftarrow rl_1^{(64)} + u^{(32)}$.

1.1.2.2. $rl_1^{(64)} \leftarrow rl_1^{(64)} + \mathrm{hi}_{(32)}\left(rl_0^{(64)}\right)$, $rl_2^{(64)} \leftarrow rl_2^{(64)} + \mathrm{hi}_{(32)}\left(rl_1^{(64)}\right)$.

1.1.2.3. $c_k^{(32)} \leftarrow \mathrm{low}_{(32)}\left(rl_0^{(64)}\right)$, $rl_0^{(64)} \leftarrow \mathrm{low}_{(32)}\left(rl_1^{(64)}\right)$, $rl_1^{(64)} \leftarrow \mathrm{low}_{(32)}\left(rl_2^{(64)}\right)$, $rl_2^{(64)} \leftarrow 0$.

1.1.3. $r_0^{(64)} \leftarrow rl_1^{(64)}$.

1.1.4. $r_1^{(64)} \leftarrow rl_2^{(64)}$.

#pragma omp section end

1.2. #pragma omp section begin

1.2.1. $rl_0^{(64)} \leftarrow 0$, $rl_1^{(64)} \leftarrow 0$, $rl_2^{(64)} \leftarrow 0$.

1.2.2. For $k \leftarrow n$, $l \leftarrow 1$, $k < nk$, $k{+}{+}$, $l{+}{+}$ do

1.2.2.1. For $i \leftarrow l$, $j \leftarrow k-l$, $i < n$, $i{+}{+}$, $j{-}{-}$ do

1.2.2.1.1. $(uv)^{(64)} \leftarrow a_i^{(32)} \cdot b_j^{32}$.

1.2.2.1.2. $rl_0^{(64)} \leftarrow rl_0^{(64)} + v^{(32)}$, $rl_1^{(64)} \leftarrow rl_1^{(64)} + u^{(32)}$.

1.2.2.2. $rl_1^{(64)} \leftarrow rl_1^{(64)} + \mathrm{hi}_{(32)}\left(rl_0^{(64)}\right)$, $rl_2^{(64)} \leftarrow rl_2^{(64)} + \mathrm{hi}_{(32)}\left(rl_1^{(64)}\right)$.

1.2.2.3. $c_k^{(32)} \leftarrow \mathrm{low}_{(32)}\left(rl_0^{(64)}\right)$, $rl_0^{(64)} \leftarrow \mathrm{low}_{(32)}\left(rl_1^{(64)}\right)$, $rl_1^{(64)} \leftarrow \mathrm{low}_{(32)}\left(rl_2^{(64)}\right)$ $rl_2^{(64)} \leftarrow 0$.

#pragma omp section end

#pragma omp parallel sections end

2. $r_0^{(64)} \leftarrow r_0^{(64)} + c_n^{(32)}$.

3. $r_1^{(64)} \leftarrow r_1^{(64)} + hi_{(32)}\left(rl_0^{(64)}\right) + c_{n+1}^{(32)}$.

4. $t^{(64)} \leftarrow hi_{(32)}\left(rl_1^{(64)}\right)$.

5. For $k \leftarrow n+2$, $k < nk$, $k++$ do

5.1. $t^{(64)} \leftarrow t^{(64)} + c_k^{(32)}$.

5.2. $c_k^{(32)} \leftarrow low_{(32)}\left(t^{(64)}\right)$.

5.3. $low_{(32)}\left(t^{(64)}\right) \leftarrow hi_{(32)}\left(t^{(64)}\right)$.

5.4. $hi_{(32)}\left(t^{(64)}\right) \leftarrow 0$.

6. $c_{nk}^{(32)} \leftarrow \mathrm{low}_{(32)}\left(r_0^{(64)}\right)$.

7. Return (c).

После выполнения работы двумя параллельными потоками п. 1.1 и 1.2 необходимо произвести корректировку пп. 2 – 6 результатов работы потока п. 1.2 посредством переноса, полученного в другом потоке п. 1.1.

На рис. 4 продемонстрируем работу алгоритма с распараллеливанием на два потока команд для $n = 3$.

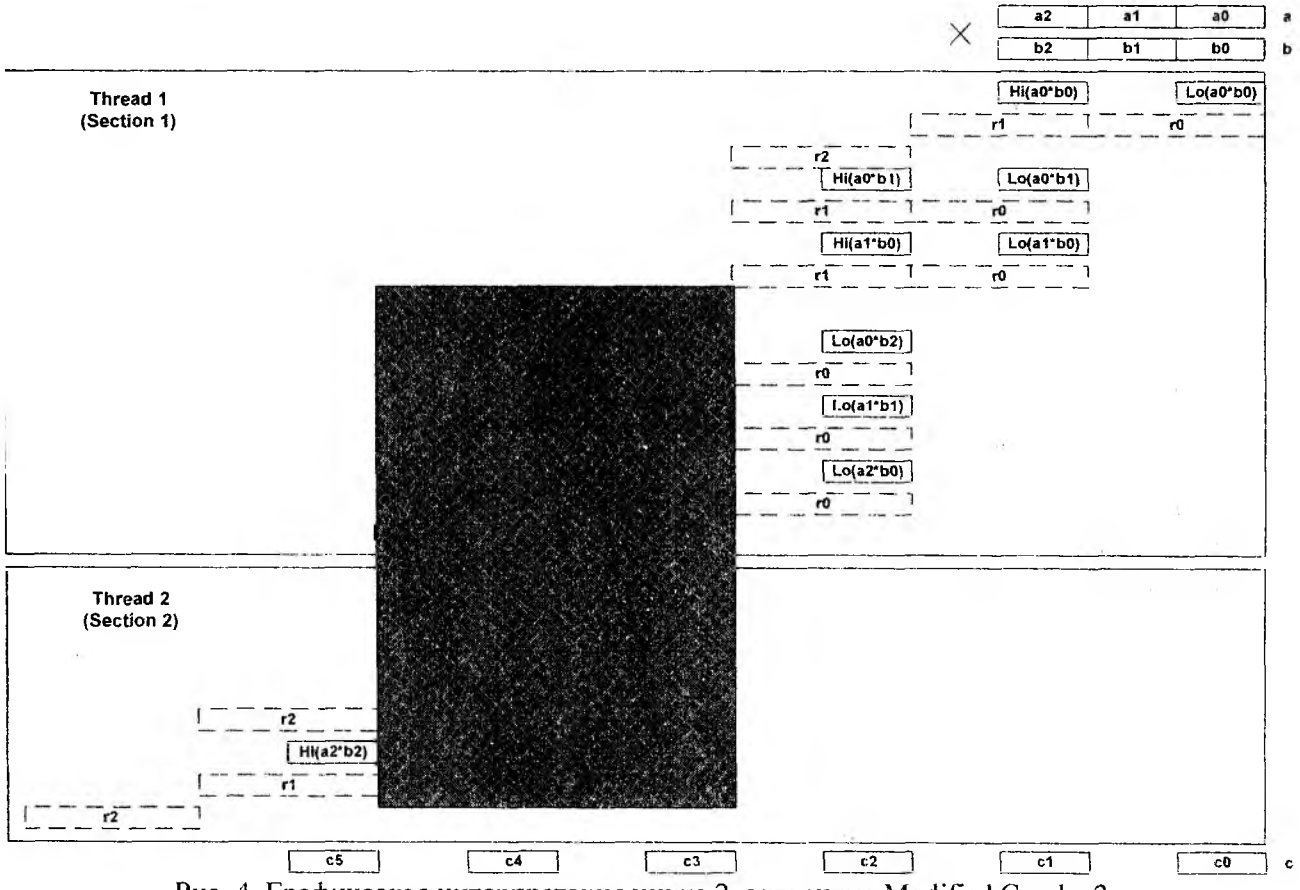


Рис. 4. Графическая интерпретация цикла 2, алгоритма Modified Comba 2x

Особого внимания заслуживает алгоритм с распараллеливанием на нескольких потоках, который изложен ниже.

Описание алгоритма Modified Comba Mx с множеством потоков

При более детальном рассмотрении алгоритма Modified Comba легко заметить, что итерации циклов п. 2 и 3 не зависят друг от друга. Исключение составляют результаты накопления сложения и переноса после текущей итерации и переходе к следующей итерации

в п. 2.2 и 2.3. Введением индивидуальных локальных переменных накопления суммы в рам ках итерации можно корректно произвести накопление суммы в итерациях цикла п.2 и 3 параллельно. Для этих целей в алгоритме Modified Comba Mx создано два массива $r0_i^{(64)}$ и $r1_i^{(64)}$, $i=\overline{0,2n-1}$. Легко заметить, что такой подход позволяет легко варьировать число параллельных потоков, без модификации алгоритма в целом.

Input: integers $a, b \in \mathrm{GF}(p)$, $w=32$, $n=\log_{2^w} a$, $nk=2n-1$.

Output: integers $c=a\cdot b$.

0. $l \leftarrow 1$.

1. #pragma omp parallel private ($r_0^{(64)}, r_1^{(64)}$) reduction (+ : l) begin

2. #pragma omp for nowait begin

2.1. For $k \leftarrow 0$, $k<n$, $k++$ do

2.1.1. $rl_0^{(64)} \leftarrow 0$, $rl_1^{(64)} \leftarrow 0$.

2.1.2. For $i \leftarrow 0$, $j \leftarrow k$, $i \le k$, $i++$, $j--$ do

2.1.2.1. $(uv)^{(64)} \leftarrow a_i^{(32)} \cdot b_j^{32}$.

2.1.2.2. $rl_0^{(64)} \leftarrow rl_0^{(64)} + v^{(32)}$, $rl_1^{(64)} \leftarrow rl_1^{(64)} + u^{(32)}$.

2.1.3. $r0_k^{(64)} \leftarrow rl_0^{(64)}$, $r1_k^{(64)} \leftarrow rl_1^{(64)}$.

#pragma omp for end

3. #pragma omp for nowait begin

3.1. For $k \leftarrow n$, $k<nk$, $k++$ do

3.1.1. $rl_0^{(64)} \leftarrow 0$, $rl_1^{(64)} \leftarrow 0$, $rl_2^{(64)} \leftarrow 0$.

3.1.2. For $i \leftarrow l$, $j \leftarrow k-l$, $i<n$, $i++$, $j--$ do

3.1.2.1. $(uv)^{(64)} \leftarrow a_i^{(32)} \cdot b_j^{32}$.

3.1.2.2. $rl_0^{(64)} \leftarrow rl_0^{(64)} + v^{(32)}$, $rl_1^{(64)} \leftarrow rl_1^{(64)} + u^{(32)}$.

3.1.3. $r0_k^{(64)} \leftarrow rl_0^{(64)}$, $r1_k^{(64)} \leftarrow rl_1^{(64)}$.

3.1.4. $l++$.

#pragma omp for end

#pragma omp parallel end

4. $rl_0^{(64)} \leftarrow r0_0^{(64)}$.

5. $rl_1^{(64)} \leftarrow r1_0^{(64)}$.

6. $c_0^{(32)} \leftarrow low_{(32)}\left(rl_0^{(64)}\right)$.

7. $rl_1^{(64)} \leftarrow rl_1^{(64)} + low_{(32)}\left(rl_0^{(64)}\right)$.

8. $rl_2^{(64)} \leftarrow hi_{(32)}\left(rl_1^{(64)}\right)$.

9. $rl_0^{(64)} \leftarrow rl_1^{(64)}$.

10. $rl_1^{(64)} \leftarrow rl_2^{(64)}$.

11. $rl_2^{(64)} \leftarrow 0$.

12. For $k \leftarrow 1$, $k<nk$, $k++$ do

12.1. $rll_0^{(64)} \leftarrow r0_k^{(64)}$.

12.2. $rll_1^{(64)} \leftarrow r1_k^{(64)}$.

12.3. $rl_0^{(64)} \leftarrow rl_0^{(64)} + low_{(32)}\left(rll_0^{(64)}\right)$.

12.4. $rl_1^{(64)} \leftarrow rl_0^{(64)} + hi_{(32)}\left(rl_0^{(64)}\right) + hi_{(32)}\left(rll_0^{(64)}\right) + low_{(32)}\left(rll_1^{(64)}\right)$.

12.5. $rl_2^{(64)} \leftarrow rl_2^{(64)} + hi_{(32)}\left(rl_1^{(64)}\right) + hi_{(32)}\left(rll_1^{(64)}\right)$.

12.6. $c_k^{(32)} \leftarrow low_{(32)}\left(rl_0^{(64)}\right)$. 12.7. $rl_0^{(64)} \leftarrow rl_1^{(64)}$. 12.8. $rl_1^{(64)} \leftarrow rl_2^{(64)}$. 12.9. $rl_2^{(64)} \leftarrow 0$.

13. $c_{nk}^{(32)} \leftarrow \mathrm{low}_{(32)}\left(rl_0^{(64)}\right)$.

14. Return (c).

Сравнение с другими алгоритмами

Эффективность распараллеливания оценивается сопоставлением среднего времени выполнения программной реализации предложенных параллельных алгоритмов с однопоточной версией алгоритма Modified Combo [11] для 1 млн. итераций. Производительность программной реализации измеряется для массивов 32-разрядных слов, что позволяет оценить производительность реализаций в целом. Предложенные модификации Modified Comba 2x, Mx и сам алгоритм Modified Comba были реализованы на C++ и скомпилированы Intel C++ Compiler XE 2011 с помощью MS Visual Studio 2005 в Release Win32 конфигурации с параметром Maximize Speed, с поддержкой SSE2. Тестирование проводилось на следующих аппаратных платформах:

- мобильных системах с CPU Intel DualCore T2130 (Microsoft Windows 7), Intel Core2 Duo T7200 (Microsoft Windows XP) и AMD A8-3510 MX (Microsoft Windows 7);
- на настольной системе среднего уровня с CPU Intel Core2 Duo E6400 (Microsoft Windows 7);
- высокопроизводительной настольной системе с CPU Intel Core i7 2600 (Microsoft Windows 7).

Все перечисленные процессоры являются двуядерными с двумя потоками выполнения команд (без поддержки Hyper Threading), в отличие от AMD A8-3510 MX (MS Windows 7), который является четырехядерным с поддержкой четырех параллельных потоков команд и Intel Core i7 2600 (MS Windows 7), который является четырехядерным с поддержкой восьми параллельных потоков выполнения команд. Результаты измерения производительности различных программных реализаций для 1 млн. умножений и вычислительных систем для заданной длины массива 32-разрядных слов приведем в таблице:

CPU	Algorithm	Count of 32-bit words / ms											
		3	4	8	16	32	48	64	96	128	192	256	384
Intel Dual Core T2130	Cmb*	16	16	46	187	702	1544	2652	5786	10246	22988	40206	90002
	Cmb* 2x	202	202	219	328	546	966	1591	3758	5381	13194	20508	49173
	Cmb* Mx	281	297	343	464	889	1544	2417	5332	8234	17935	33203	72519
Intel Core2 Duo T7200	Cmb*	16	16	46	156	484	1015	1734	3766	6719	14703	26063	57735
	Cmb* 2x	234	172	187	235	422	703	1125	2157	3641	7812	13531	29859
	Cmb* Mx	266	281	297	406	719	1235	1844	3593	6015	12891	22500	49625
Intel Core2 Duo E6400	Cmb*	0	16	31	94	328	688	1172	2515	4500	9843	17438	38610
	Cmb* 2x	78	93	93	125	266	453	719	1672	2438	5265	9547	20281
	Cmb* Mx	141	141	172	250	453	781	1218	2735	4047	8688	16000	34578
AMD A83510	Cmb*	16	16	31	156	452	921	1576	3682	6537	14212	24133	51480
	Cmb* 2x	187	187	209	250	421	687	1061	2075	3416	7472	13072	28205

MX	Cmb* Mx	359	375	390	437	593	826	1185	2075	3276	6755	11404	24804
Intel Core i7-2600	Cmb*	0	16	16	78	249	546	936	2044	3525	5554	13884	30872
	Cmb* 2x	124	109	109	156	266	484	764	1622	2605	3635	9734	21902
	Cmb* Mx	172	156	203	234	312	421	609	1139	1794	7831	6334	13089

Проанализируем результаты экспериментов, приведенные в таблице. Так, на платформе с процессором Dual Core T2130 над однопоточной реализацией Cmb* свое превосходство показал Cmb* 2x на числах длиной в 32 слова, и Cmb* Mx на числах длиной в 48 слов. Отметим, что Cmb* Mx показал худшую производительность, чем Cmb* 2x. Процессор Core2 Duo T7200 оказался значительно производительнее Dual Core T2130, в связи с этим однопоточную реализацию Cmb* смогли превзойти Cmb* 2x на числах длиной в 32 слова, а Cmb* Mx на числах длиной в 96 слов. Как и в предыдущем случае, Cmb* Mx показал значител ьно худшую производительность, чем Cmb* 2x. Такое поведение легко объясняется более высокой производительностью ядер Core2 Duo T7200 в сравнении с Dual Core T2130.

Особого внимания заслуживает мобильный процессор AMD A83510 MX, который показал превосходство Cmb* 2x на числах длиной в 32 слова, а Cmb* Mx на числах длиной в 48 слов, кроме того, реализация Cmb* Mx превзошла Cmb* 2x на числах длиной в 96 слов. Данные результаты позволяют заявить, что Cmb* Mx может эффективно применяться на многопроцессорных/многоядерных вычислительных системах.

Перейдем к рассмотрению результатов экспериментов на настольных системах.

Система Core2 Duo E6400 показала схожее поведение с Core2 Duo T7200: над однопоточной реализацией Cmb* свое превосходство показал Cmb* 2x на числах длиной в 32 слова и Cmb* Mx на числах длиной в 128 слов. Такое поведение также объясняется более высокой производительностью ядер Core2 Duo E6400, в сравнении с Core2 Duo T7200.

Наилучшие показатели производительности оказались у системы с процессором Core i7-2600. Так, благодаря высокой производительности ядер превосходство Cmb* 2x и Cmb* Mx над Cmb* проявилось лишь на числах длиной в 48 слов. При этом эффект от выполнения Cmb* Mx на четырех ядрах в восьми потоках проявился в явном превосходстве не только над Cmb*, но и над Cmb* 2x. Дальнейшее увеличение длины чисел показало значительное превосходство Cmb* Mx над Cmb* 2x.

Из таблицы видно, что эффект от распараллеливания начинает проявляться на числах длиной более 32 слов. Это связано с достаточно большими накладными расходами, которые вызваны созданием нового потока в рамках одной операции умножения, с учетом того, что эти накладные расходы соизмеримы с расходами на само умножение. Избавиться от такого нежелательного эффекта можно, если параллельные потоки создавать заранее, перед выполнением всех арифметических операций, на этапе инициализации всей библиотеки.

Также следует обратить внимание на другой эффект: чем выше производительность одного ядра, тем на большей длине числа проявляется эффект от распараллеливания, об этом свидетельствуют результаты измерений на процессорах Intel Core i7-2600 и AMD A83510 MX.

Выводы

- Предложенный авторами алгоритм Modified Comba может быть эффективно распараллелен Так, на процессоре Intel Core i7-2600 удалось добиться превосходства в 1,5 раза для Modified Comba 2x и в более чем два раза – для Modified Comba Mx, в сравнении с однопоточной реализацией.
- Эффект от распараллеливания начинает проявляться при длине в 32 слова по 32 бита (1024 бита), что говорит о значительных накладных расходах на создание параллельного потока, компенсацию которых следует закладывать непосредственно в саму библиотеку арифметических преобразований на этапе инициализации.
- Проведенные исследования показали, что компиляторы GNU gcc C++ (Linux Debian 6.0 x64) и MS Visual Studio C++ (Windows XP, Windows 7) показали себя значительно хуже,

ием Intel C++ Compiler XE 2011 (Windows 7), т.к. скомпилированные ими программные реализации показали значительно худшую производительность, поэтому не приводятся в работе. Худшая производительность обусловлена большим временем, которое тратится программной реализацией на создание параллельного потока, и временем задержки перед уничтожением потока после его отработки.

- В качестве дальнейшего направления работы видится применение подхода распараллеливания к другим алгоритмам арифметических операций в кольцах и полях для повышения производительности криптографических операций с отрытым ключом, например на алгебраических кривых. О необходимости проведения данных исследований говорят результаты работы [12] по использованию технологии CUDA при реализации криптосистемы на эллиптических кривых.

Список литературы: 1. *Diffie, W., Hellman, M. E.* New directions in cryptography // IEEE Transactions on Information Theory. vol. IT-22, pp. 644–654, 1976. 2. *Selçuk Baktir and Erkay Sava.* Highly-Parallel Montgomery Multiplication for Multi-core General-Purpose Microprocessors // Cryptology ePrint Archive. - Report 2012/140. –2012. – 16 p. Available at: http://eprint.iacr.org. 3. *Pascal Giorgi, Thomas Izard, Arnaud Tisserand.* Comparison of Modular Arithmetic Algorithms on GPUs // In Proc. International Conference on Parallel Computing (ParCo 2009). – Vol19. – Lyon, France. -2009. – pp. 315-322. 4. *The OpenMP* API Specification for Parallel Programming. Available at: http://openmp.org. 5. *OpenMP* in Visual C++. Available at: http://msdn.microsoft.com/en-us/library/tt15eb9t.aspx. 6. Khronos OpenCL API Registry. URL: http://www.khronos.org/registry/cl/. 7. Intel Threading building blocks for open source. URL: http://threadingbuildingblocks.org/. 7. *NVIDIA* CUDA. URL: http://www.nvidia.ru/object/cudahome new_ru.html. 8. AMD Accelerated Parallel Processing (APP). URL: http://developer.amd.com/sdks/AMDAPPSDK/samples/showcase/Pages/default.aspx. 9. *Comba, P. G.* Exponentiation cryptosystems on the IBM PC // IBM Systems Journal. – Vol. 29(4). – 1990. – pp. 526–538. 10. *Kovtun, V., Okhrimenko, A.* Approaches for the performance increasing of software implementation of integer multiplication in prime fields // Cryptology ePrint Archive. – Report 2012/170. – 2012. – 9 p. Available at: http://eprint.iacr.org. 11. *Giorgi P. Izard T, Tisserand A.* Comparison of Modular Arithmetic Algorithms on GPUs. URL: http://hal-lirmm.ccsd.cnrs.fr/lirmm-00424288/fr/

Национальный авиационный университет, Киев

Поступила в редколлегию 13.09.2012

УДК 004.056.51

О.А. ШЕВЧУК, Ю.І. ГОРБЕНКО, канд. техн. наук

АВТЕНТИФІКАЦІЯ АПАРАТНИХ ЗАСОБІВ КЗІ

Розвиток ринку засобів криптографічного захисту інформації (КЗІ) надає актуальності як новим задачам, так і стимулює розвиток підходів до існуючих, таких як задача побудування моделей порушника та загроз. Зростання числа сторін, що беруть участь у виробництві засобів КЗІ, поставці, інсталяції, супроводженні в експлуатації, використанні тощо, вимагає додаткового аналізу вразливостей компонент системи в змінених умовах [7].

В [1] було розглянуто вразливість схем ЕЦП, що ґрунтуються на перетвореннях в групі точок еліптичної кривої, до зловживань з боку виробника.

Метою статті є аналіз впливу визначеної загрози з врахуванням деякої третьої сторони, що може бути задіяна в процесі розповсюдження апаратних засобів, а також розробка рекомендацій.

Огляд існуючих вимог та рекомендацій до апаратних засобів КЗІ

В світових стандартах та рекомендаціях [4, 5] існує ряд вимог щодо виконання процесів на усіх етапах життєвого циклу криптографічних засобів. Розглянемо узагальнене тлумачення можливостей порушника з урахуванням виконання визначених процедур:

- порушника відокремлено від процесу розробки та створення засобу КЗІ;
- порушник в певній мірі має фізичний доступу до засобу КЗІ ще до закінченя процесу інтеграції та безпосереднього використання засобу користувачем;
- компрометація ключової інформації за рахунок прямого доступу до елементів збереження ключової інформації неможлива, так як засіб КЗІ ще не пройшов етапу персоналізації;
- компрометація ключової інформації або ключів сесій та векторів ініціалізації за рахунок аналізу показників електромагінтного випромінювання, живлення та інших фізичних каналів витоку не має сенсу, так як засіб КЗІ ще не пройшов етапу персоналізації.
- не існує можливість зміни криптографічних алгоритмів за рахунок прямого доступу до елементів збереження алгоритмів перетворень або фізичної модифікації пристрою.

Також будемо вважати, що [1, 4]:

- порушника відокремлено від процесу та середовища використання засобу КЗІ;
- порушник не може перехопити та втручатися у канал зв'язку з засобом;
- порушник не може використовувати фізичні канали витоку інформації для персоналізованого засобу КЗІ;
- порушник може використати результати некоректної роботи модуля, що викликані недоліками реалізації та не виявлені в процедурі самотестування, якщо вони знаходяться у відкритому доступі;
- недійсні (з помилками) результати криптографічних перетворень можуть знаходитися у відкритому доступі;
- результати криптографічних перетворень, що були отримані у помилковому стані, можуть знаходитися у відкритому доступі та призвести до компрометації ключової інформації;
- використання засобу КЗІ у відкритому середовищі не розглядається.

Введені вище моделі порушника та загроз, які він може реалізувати, дозволяють запропонувати заходи протидії загроза порушника. Проведений аналіз показав, що їх необхідно розробляти перше за все з урахуванням вимог міжнародного стандарту FIPS-140 [5] та стандарту PCI HSM [4]. Основними з них можуть бути такі:

- наявність та застосування самотестування модуля КЗІ;
- повна перевірка справжності, цілісності та автентичності даних оновлення;

- автентифікація користувача, та впровадження розподілу доступу;
- впровадження відповідного рівня забезпечення фізичної цілісності засобу КЗІ.
- контроль фізичної цілісності модуля криптографічного захисту.

Наведені рекомендації та оцінки дозволяють зробити висновок, що у випадку дотримання рекомендацій порушник не може внести не випадкові (детерміновані) зміни до засобу (модуля КЗІ). Але в [4, 5] та і наведених вище рекомендаціях не визначено, які заходи потрібно прийняти для протидії загрозам, що пов'язані з підміною засобів КЗІ. Тобто, користувач не може однозначно визначити, чи є отриманий засіб КЗІ саме тим виробом, з характеристиками якого він ознайомлений чи навіть закупив його. Розглянемо вразливості такої загрози, зважаючи на необхідність перекриття такої загрози, розглянемо та проведемо більш детальний аналіз загрози підміни засобу КЗІ, але за умови використання електронного цифрового підпису (ЕЦП).

Аналіз та оцінка вразливості загрози підміни засобу КЗІ

Із аналізу [1, 4] можна зробити висновок, що загроза підміни засобу КЗІ, що надає послугу ЕЦП у визначеному середовищі, може бути реалізована за таких умов:

- порушник не може модифікувати інтерфейс обміну з засобом КЗІ для компрометації ключових даних, так як він не має доступу до засобу під час та після персоналізації;
- порушник не може створити додатковий фізичний канал витоку ключової інформації;
- порушник може створити додатковий канал витоку ключової інформації з результатом обробки вхідних даних;
- порушник може ініціювати підпис довільних даних, та додати результат обчислень до дійсного результату;
- порушник може замінити процедуру випадкової генерації особистого ключа та ключів сесії на псевдовипадкову, з відомим вектором ініціалізації.

Аналіз ряду джерел та проведені дослідження [1] дозволяють уточнити умови їх реалізації та запропонувати заходи протидії. Оскільки порушник не може взаємодіяти з засобом КЗІ під час та після персоналізації модуля, то єдиним каналом обміну інформації є відкриті документи, що з використанням відповідного засобу КЗІ.

Далі, для захисту програмного засобу КЗІ від загрози підміни необхідно та достатньо забезпечити цілісність та автентичність програмного забезпечення, що використовується. Дійсно, нехай існує програмна реалізація криптографічного алгоритму $S(k,m)$, що представлена у вигляді повідомлення $M_{S(k,m)}$, та його цифрового підпису $S(k,M_{S(k,m)})$. Нехай також порушник може створити новий алгоритм $S'(k,m)=\alpha\cdot S(k,m)$, що відрізняється від оригінального, наприклад у вигляді повідомлення $M_{\alpha\cdot S(k,m)}$. Нехай також $S(k,M_{\alpha\cdot S(k,m)})=S(k,S(k,m))$. Тоді $M_{\alpha\cdot S(k,m)}=M_{S(k,m)}$, та відповідно $\alpha\cdot S(k,m)=S(k,m)$. Тоді $\forall k\in K, m\in M:\alpha\cdot S(k,m)=S(k,m)$. Але $\alpha\cdot S(k,m)\neq S(k,m)$.

Далі, у разі виконання засобу КЗІ як програмно-апаратного засобу, або апаратного, користувач не має можливості представити алгоритм його роботи у вигляді тексту, та упевнитися у цілісності та автентичності алгоритмів, що реалізує засіб.

Дійсно, для дослідження справжнього алгоритму, за яким виконує обчислення апаратний засіб, необхідна фізична декомпозиція корпусу захисту засобу та його внутрішніх компонентів, що призведе до припинення функціонування пристрою.

Розглянемо випадок, коли відсутнє засвідчення справжності(дійсності) засобу КЗІ, що надає послуги ЕЦП в визначеному середовищі. Для міжнародних стандартів цифрових підписів, в тому числі і українського стандарту ДСТУ 4145:2002, використовуються такі елементи, що формуються випадковим чином:

- сеансовий ключ;
- особистий ключ.

До них висуваються такі вимоги:

- випадковість.
- контроль цілісності;
- конфіденційність;
- непередбачуваність;
- необоротність.

Проведемо аналіз можливих впливів порушника на забезпечення визначених вимог т наслідки цього впливу. По-перше, особистий ключ X_a обчислюється за допомогою деяко випадкової функції RNG(). Властивості, алгоритми та вимоги до випадкової функції RNG(), що використовуються під час процесу генерації ключових даних, визначено стандартами FIPS 186-3, NIST SP 800-50 та іншими. Особистий ключ обчислюється під час персоналізації пристрою, та у разі необхідності, в тому числі в процесі використання. Але порушник може виконати деяку модифікацію псевдовипадкової функції, як $\alpha \cdot \mathrm{RNG}()$, для зміни властивості непередбачуваності, але із зберіганням властивості випадковості, наприклад утворюючи генератор псевдовипадкових послідовностей з відомим проміжком стартових значень $log_2 \#\{s_1, s_2, \ldots s_n\} < \Theta$.

У цьому випадку на етапі персоналізації значення особистого ключа користувача засобу КЗІ набуде таких значень, що порушник може обчислити його за поліноміальний час. Далі, за наявності послуг імпорту та експорту ключових даних, користувач може генерувати ключ на довіреному засобі, та завантажити його в засіб. В такий спосіб користувач може контролювати проходження персоналізації.

Подальший розгляд пов'язаний з аналізом можливих варіантів персоналізації криптографічного засобу, що модифікований визначеним шляхом. Нехай порушник виконав підмінений (компрометований) апаратний засіб таким чином, що в в процесі персоналізації використовується ключ, який може бути обчислений порушником за поліноміальний час. Нехай також користувач створює деякий и особистий ключ X_a, та відповідний йому відкритий ключ P_a за допомогою уже компрометованого засобу. Тоді, $X_a = \alpha \cdot RNG()$, та відповідний йому відкритий ключ $P_a = GenPubKey(X_a, \ldots)$, де $GenPubKey$ – деяка функція обчислення відкритого ключа. Користувач за допомогою засобу обчислює деякий цифровий підпис $Sign(X_a, m)$, та, не використовуючи, засіб перевіряє цифровий підпис $Verify(P_a, Sign(X_a, m))$. Підпис буде дійсним, так як він відповідає відкритому ключу P_a, а його було створено за допомогою засобу уже компрометовного ключа X_a.

Далі, нехай порушник використав підмінений (компрометований) апаратний засіб таким чином, що в в процесі персоналізації використовується ключ, який може бути обчислений порушником за поліноміальний час. Нехай також користувач завантажує деякий особистий ключ X_a, та відповідний йому відкритий ключ P_a, якого було отримано за стороннім довіреним генератором. За такої умови ключ X_a відповідає вимогам, та є непередбачуваним для порушника. Нехай апаратний засіб обчислює деякий компрометований особистий ключ $\alpha(X_a)$ та користувач за допомогою засобу КЗІ обчислює деякий цифровий підпис $\mathrm{Sign}(X_a, m)$ Далі користувач, не застосовуючи засіб КЗІ, перевіряє цифровий підпис $\mathrm{Verify}(P_a, \mathrm{Sign}(X_a, m))$ Якщо $\alpha(X_a) = X_a$, тоді перевірка дійсна, причому особистий ключ відповідає тому, що був завантажено користувачем. В іншому випадку користувач констатує факт компрометації ключа.

В той же час порушник може використати особливості алгоритмів ЕЦП для компрометації особистого ключа, який було отримано за довіреним каналом, та завантаженс у компрометовний пристрій.

Наведена в [1] вразливість дозволяє порушнику компрометувати особистий ключ із наступними загальними характеристиками:

• за відношенням до АС – зовнішній;
• за місцем дії – з відсутнім доступом до контрольованої зони;
• за часом дії – до персоналізації засобу КЗІ;
• за рівнем та можливостями: порушник I-II рівня, з можливістю замовити розробку засобу КЗІ за власним технічним завданням
• порушнику відома загальна структура, функції та алгоритми взаємодії з засобом КЗІ.

Як було позначено, у діючих міжнародних нормативних документах не виділяють цей клас атак, але в вимогах до апаратних засобів КЗІ, що визначені в Payment Card Industry (PCI), сформовані адміністративні вимоги, виконання яких спрямовано на запобігання реалізації визначених загроз.

В процесі досліджень встановлено, що специфікація PCI HSM 1.0 від 2009 р. [8] визначає основоположні адміністративні вимоги до створення та транспортування засобів КЗІ через сітку дистрибуції.

Сутність вимог зводиться до наступного:

• дизайн криптографічного засобу має захищати пристрій від підміни таким чином, щоб створення подібного було економічно невигідно;
• якщо криптографічний засіб має бути автентифіковано, в розумінні наявності унікального секрету пристрою, тоді цей секрет має бути унікальним для кожного криптографічного засобу, невідомий та непередбачуваний довільній особі, та завантажений до пристрою з подвійним контролем, для запобігання компрометації секрету;
• криптографічний засіб має бути доставлено до місця персоналізації за визначеним маршрутом, що передбачає аудит та контроль місце знаходження кожного пристрою в кожний проміжок часу;
• обов'язкова явна процедура передачі пристрою від виробника до особи, що здійснює персоналізацію;
• за час пересування від виробника до особи, що здійснює персоналізацію, пристрій має знаходитись у пакуванні, що дозволяє виявити неавторизований доступ до пристрою.

Далі, в новій специфікації PCI HSM 2.0[4], що набуде сили в 2012 р., внесено ряд додаткових вимог, в тому числі:

• виробник має надавати інструкції, в тому числі в глобальній мережі, щодо перевірки користувачем цілісності та справжності засобу, якщо це неможливо, необхідно виконання умов (D1) з PCI HSM 1.0 [8];
• якщо в доставці пристрою беруть участь декілька сторін, тоді перевірка пакування та вимог є обов'язком кожної сторони;
• документація до пристрою має містити інструкції щодо автентифікації критичних компонентів системи;
• якщо засіб КЗІ персоналізується виробником, то виробник має перевірити автентичність засобу перед завантаженням;
• якщо засіб КЗІ персоналізується іншою стороною, виробник має надати можливість перевірити автентичність засобу перед персоналізацією;
• кожен засіб має мати унікальний ідентифікатор, або має бути ідентифікований криптографічним шляхом.

Необхідно зазначити, що PCI HSM [4,8] не вказує на методи забезпечення визначених вимог. На нашу думку, використання криптографічних алгоритмів та примусова автентифікація засобу користувачем на етапах персоналізації є суттєвим та невід'ємним заходом.

Проведені дослідження та отримані результати дозволяють зробити наступні висновки та рекомендації:

• засоби КЗІ, що надають послуги цифрового підпису особливо вразливі до загрози підміни.

• за умов підміни пристрою, та публікації документів з цифровим підписом в відкритих джерелах, що обчислено за допомогою засобу, порушник може компрометувати особистий ключ користувача з поліноміальною складністю;

• задача визначення наявності порушення вимог до генерації ключа сеансу є обчислювальне складною.

• проблема автентифікації засобів КЗІ визнана актуальною, що підтверджують зміни до адміністративних вимог процесів забезпечення інформаційної безпеки, зокрема PCI HSM [4, 8].

• нині не опубліковано рекомендацій та не прийнято стандартів та регламентів, що визначають протоколи та методи надання послуги автентифікації засобів КЗІ.

• на ринку криптографічних засобів України не представлено засобів, що надають послугу автентифікації;

• доцільно розробити вимоги та рекомендації щодо реалізації послуги автентифікації апаратного засобу КЗІ;

• в рамках розробки вимог необхідно визначити криптографічні методи автентифікації.

Метод автентифікації засобу криптографічного захисту інформації, вразливого до підміни

Автентифікація являє собою процес встановлення достовірності твердження, що суб'єкт або об'єкт має очікувані властивості. По суті, механізм дозволяє одній сутності перевірити визначені властивості іншої, а реалізацію механізму автентифікації можна розглядати як метод захисту користувачів та власників інформація та ресурсів від різних видів обману. Механізми автентифікації реалізуються на основі застосування криптографічних протоколів та дозволяють захистити інформацію чи ресурси від їх модифікації, підміни чи створення хибних. Аналіз показав, що механізми автентифікації та їх реалізації у вигляд криптографічних протоколів дозволяють здійснити автентифікацію джерела, в тому числі:

• перевірку цілісності джерела повідомлення;
• ідентифікацію відправника повідомлення;
• перевірку цілісності отриманих даних;
• перевірку актуальності повідомлення;
• перевірку дійсності відправника.

Відносно автентифікації сутності може здійснюватись:

• автентифікація клієнта;
• автентифікація сервера;
• автентифікація засобу взагалі;
• одностороння чи взаємна автентифікація;
• генерацію автентифікованих ключів.

Враховуючи наведене, розглянемо передумови до необхідності виконання протоколу автентифікації засобу КЗІ. Із наведеного випливає, що автентифікація пристрою зумовлена наявністю вразливостей відносно апаратних реалізацій алгоритмів ЕЦП, що засновано на схемах DSA Ніберг – Рюпеля, в першу чергу до загрози підміни. Вказане можна пояснити наступним:

• реалізація загрози на будь-якому етапі циклу використання засобу може призвести до компрометації особистого ключа користувача;

• виробник апаратного засобу є стороною, якій довіряє користувач;

• виробник гарантує якість засобу КЗІ;

• компрометація засобу не має призводити до компрометації інших засобів;

• виробник не може мати відомості щодо усіх можливих користувачів засобу, але має мати відомості про усі пристрої, що було створено;

• засіб КЗІ та виробник не мають можливості оновлювати ключову інформацію;

• користувач до персоналізації засобу може не мати можливості його використання.

У цілому, згідно з наведеною моделлю загроз, необхідно зробити висновок про необхідність односторонньої автентифікації сутності (засобу КЗІ). Це пояснюється тим, що в рамках визначеної моделі загроз, критичним до загрози підміни є проміжок часу між створенням засобу виробником та персоналізацією засобу користувачем. Таким чином, процес автентифікації пристрою не пов'язаний з процесом автентифікації та авторизації користувача пристроєм після персоналізації, та має розглядатися відокремлено. Проведений аналіз [1, 4] показав, що в указаних стандартах не наведені вимоги до автентифікації сутності, але є посилання на протоколи узгодження ключів ISO/IEC 11770-2, ISO IEC 11770-3. Комітетом міжнародної стандартизації ISO/IEC узгоджено групу стандартів ISO/IEC 9798, що містять алгоритми автентифікації та посилаються на ISO/IEC 11770-2,3. Розглянемо протокол автентифікації стандарту ISO IEC 9798-2,№6.1, що є еквівалентним протоколу №8, наведеному в ISO/IEC 11770-2.

Сутність протоколу є у наступному:

• ініціатор перевірки U відсилає третій довіреній стороні згенерований новий параметр N_u та розпізнавальний ідентифікатор D.

• третя довірена сторона V створює та відсилає об'єкту U маркер, що містить параметри новизни N_u та ключ сенсу K_{UD}, ідентифікатор D, що зашифровані загальним з V ключем K_{VU}. Також V створює та відсилає новий параметр новизни N_v, сессійний ключ K_{UD} та ідентифікатор U, що зашифровані загальним з V та D ключем K_{VD};

• об'єкт D перевіряє першу частину маркеру шляхом розшифрування його зашифрованої частини спільним ключем K_{VD} та перевіряє: дійсність параметру новизни; дійсність ідентифікатору D.

Після цього U формує та пересилає об'єкту D другу частину отриманого маркеру, створює та відсилає маркер, що зашифрований спільним ключем U та D – K_{UD}, випадковий параметр N_{UD} та ідентифікатор B.

• Об'єкт D, отримавши маркери, розшифровує першу частину спільним ключем K_{VD}, перевіряє дійсність ідентифікатору U та відновлює спільний секрет K_{UD}. За спільним секретом K_{UD} об'єкт D розшифровує та перевіряє дійсність ідентифікатору D у другому маркері.

• Далі D відсилає U зашифрований спільним ключем K_{UD} маркер, що містить ідентифікатор U та випадковий параметр N_{VD}.

• U розшифровує маркер спільним ключем K_{VD} та перевіряє дійсність ідентифікатору U. У разі успішного виконання автентифікація завершена.

Повний протокол:

1	U	→	V:	N_U, D
2	V	→	U:	$\{N_U, K_{UD}, D\}_{K_{VU}}, \{N_V, K_{UD}, U\}_{K_{VD}}$
3	U	→	D:	$\{N_V, K_{UD}, U\}_{K_{VD}}, \{N_{UD}, D\}_{K_{UD}}$
4	D	→	U:	$\{N_{UD}, U\}_{K_{UD}}$
5			U:	$U^D = U^U$

Наведений вище криптографічний протокол має такі характеристики:

• для процесу автентифікації в протоколі використовується спільний для двох об'єктів секретний ключ, що розподілено між сторонами;

• один з об'єктів дає запит третій довіреній стороні на отримання ключа K_{UD};

• новизна забезпечується протоколом використанням змінного у часі параметру N_t та розподіленим ключем K_{UD};

• сторони U та D є рівноправними, причому обидві сторони розділяють ключі K_U та K_D з третьою довіреною стороною;

• протокол має суттєвий недолік, так як адміністративним стандартом PCI HSM 2.0[4] визначено, що можливість перевірка автентичності пристрою має бути надана особі, що здійснює персоналізацію пристрою, а також сторонам, що виконують доставку пристрою.

Аналіз наведених характеристик дозволяє висунути до всіх сторін, що задіяні в транспортуванні, зберіганні, реалізації та персоналізації пристрою, такі вимоги:

• їм необхідно звернутися до виробника для вироблення ключів;

• оновлювати ключі у випадку компрометації, або через визначений проміжок часу.

• виробник зобов'язаний зберігати інформацію про всіх постачальників (в тому числі і проміжних) та клієнтів;

• сторони U та D мають знати про існування одна одної, що викликає істотні труднощі ще на етапі створення засобу.

Приклад розподілу секрету наведено у таблиці:

Розподіл секрету у протоколі ISO/IEC 9798-2.6.1 – (*а*) та запропонованому – (*б*)

→	V	U	D
V	---	√	√
U	√	---	X
D	√	X	---
а			

→	V	U	D
V	---	X	√
U	√	---	X
D	√	X	---
б			

Пропонується зробити модифікацію розподілу секрету (таблиця (*а*)), змінивши симетричні криптоперетворення для обміну таким чином:

$$1 \quad U \rightarrow V: \quad \{D, T_{CT}\}_{PK(V)}$$
$$2 \quad V \rightarrow U: \quad \{\{N_V, T_{VD}\}_{K_{VD}}, \{N_V\}_{T_{VD}}\}_{T_{UV}}$$

Удосконалений протокол має наступні властивості:

• спрощується задача зберігання секретів виробником, при цьому виробник не має турбуватися про факти компрометації секрету користувача, оновлення секрету та ідентифікацію;

• забезпечується захист від перехопленням порушником токену перевірки $\{N_V\}_{T_{VD}}\}_{T_{UV}}$.

Вказане необхідне для унеможливлення атаки за паралельними каналами, у випадку використання мережевого засобу КЗІ.

Якщо засіб КЗІ не має доступу до мережі, тоді перші кроки протоколу можуть бути спрощені наступним чином:

$$1 \quad U \rightarrow V: \quad D, N_{UV}$$
$$2 \quad V \rightarrow U: \quad \{\{N_V, T_{VD}\}_{K_{VD}}, \{N_V\}_{T_{VD}}, N_{UV}\}_{SK(V)}$$

Це можливо, так як компрометація токену не призведе до здійснення атаки паралельними каналами.

З урахуванням наведеного остаточний варіант запропонованого криптографічного протоколу автентифікації має наступний вигляд:

$$1 \quad U \rightarrow V: \quad \{D, T_{UV}\}_{PK(V)}$$
$$2 \quad V \rightarrow U: \quad \{\{N_V, T_{VD}\}_{K_{VD}}, \{N_V\}_{T_{VD}}\}_{T_{UV}}$$
$$3 \quad U \rightarrow D: \quad \{N_V, T_{VD}\}_{K_{VD}}$$

4 D → U: $\{N_1\}^D_{T_{UD}}$

U: $\{N_1\}^D_{T_{UD}} = \{N_1\}^C_{T_{UD}}$

Висновки

При створенні засобів КЗІ та їх використанні необхідно враховувати, що для апаратних та програмних засобів КЗІ повинно розроблятися різні моделі загроз;

1.Для схем ЕЦП, що ґрунтуються на алгоритмах DSA та Ніберг – Рюпеля, існує реальна загроза повного розкриття для випадку реалізації вразливості до лінійно пов'язаних ключів сеансу.

2.Діючі світові стандарти та нормативні документи, такі як FIPS-140 та PCI HSM не відокремлюють загрозу повного розкриття через реалізацію загрози підміни. Рекомендації PCI HSM визначають тільки ряд адміністративних заходів, при виконанні яких забезпечується можливість протидії реалізації загрози підпміни, але тільки на проміжку між створенням засобу КЗІ та його персоналізацією.

3.PCI HSM визначає можливість автентифікації засобу КЗІ особою, що здійснює персоналізацію, але не визначає посилання та інші документи або методи, за якими вона має бути здійснена.

4.Запропоновані модифікація механізму та криптографічного протоколу ISO/IEC 9798-2 № 6.1. дозволяє усунути визначені недоліки.

- реалізуючи лише алгоритми симетричного шифрування у засобі КЗІ;
- забезпечуючи виконання вимог PCI HSM 2.0 щодо доставки засобів КЗІ за допомогою третіх служб без попередньої реєстрації та розподілу ключового матеріалу та обов'язкової реєстрації користувачів.

Список літератури: 1. *Шевчук, О.* Актуальність атаки на зв'язанних ключах для апаратних реалізацій засобів КЗІ // Радиотехника. – 2011. – № 166. – С. 70 – 75. 2. *Шевчук, О.* Алгоритми відновлення повідомлення ЕЦП стандарту ISO/IEC 9796-3 та екзистенціальна підробка підписів, що засновується на них // Прикладная радиоэлектроника. – 2011. – Т. 10, № 2. – С. 183-187. 3. *Gallagher, P., Foreword, D. D., Director, C. F.* FIPS PUB 186-3 FEDERAL INFORMATION PROCESSING STANDARDS PUBLICATION Digital Signature Standard (DSS). – 2009. 4. *Hardware* Security Module (HSM) v2.0. – 2012. – Secuirty Standards Council. – URL: https://www.pcisecuritystandards.org/security_standards/documents.php. 5. Terryn W. Fips 140-3. – International Book Marketing Service Limited, 2011. – URL: http://books.google.com.ua/books?id=fjBWXwAACAAJ. 6. *Gallo, R., Kawakami, H., Dahab, R.* On device identity establishment and veriication // inproceedings of the 6th European conference on Public key infrastructures, services and applications. – Pisa, Italy: Springer-Verlag, 2010. – С. 130 – 145. – (EuroPKI'09). – URL: http://dl.acm.org/citation.cfm?id=1927830.1927843.0 7. FIPS 140-1 and FIPS 140-2 Vendor List. – URL:http://csrc.nist.gov/groups/STM/cmvp/documents/140-1/1401vend.htm. 8. Hardware Security Module (HSM) v1.0. – 2009. – Secuirty Standards Council. – URL:https://www. Pcisecuritystandards.org/documents/PCI HSM Security Requirements v1.0 final.pdf.

Харківський національний університет радіоелектроніки

Надійшла до редколегії 17.09.2012

УДК 681.3.06

О.Г. ХАЛИМОВ, А.Н. ГЕРЦОГ

УНИВЕРСАЛЬНОЕ ХЕШИРОВАНИЕ ПО ОБОБЩЕННЫМ КРИВЫМ ГУРВИЦА

Универсальное хеширование по алгебраическим кривым χ над конечным полем F_q на основе скалярного произведения по рациональным функциям линейного базисного пространства $f_i \in F_q(\chi) \setminus \{0\}$ для сообщения $m = (m_1,...,m_k)$, $m_i \in F_q$ в точке кривой P_j определяется вычислением $h_{P_j}(m) = \sum_{i=1}^{k} f_i(P_j) m_i$ и имеет оценку для вероятности коллизии $\varepsilon = \rho_k / N$, где ρ_k – максимальное значение полюса рациональной функции f_k и N – число точек алгебраической кривой [1]. Универсальное хеширование $h_{P_j}(m)$ обеспечивает наилучшие соотношения между вероятностью коллизии, затратами на ключевые данные, сложность вычисления при фиксированной длине сообщения и поля вычисления. Основная проблематика реализации универсального хеширования по алгебраическим кривым состоит в построении проективного многообразия с большим отношением числа точек кривой к её роду. Универсальное хеширование по кривым Гурвица реализует вычисления на большом числе точек. Для случая простого поля по сравнению с кривыми Ферма достигается абсолютно наилучший результат [2]. Основные результаты по кривым Гурвица представлены в работах [3 – 5]. В работе [3] представлено решение задачи построения максимальных кривых Гурвица. Условия существования и построения нетривиальных кривых Гурвица получены в [4, 5]. В работе [5] впервые определены все семейства нетривиальных кривых Гурвица. Оценки параметров наилучших кривых Гурвица в простом, кубическом поле представлены в [2, 6]. Здесь же рассмотрены частные случаи кривых Гурвица с большим числом точек. Важной научной задачей является разработка метода построения хороших кривых Гурвица без ограничений на показатели степени кривой над произвольным конечным полем с уменьшенной сложностью вычислений.

В данной работе предлагается метод построения кривых Гурвица с наибольшим отношением числа точек к роду на основе приведения к обобщенным кривым минимального рода с перебором значения наибольшего показателя кривой. С этой целью в разд. 1 приводятся определение и основные результаты по кривым Гурвица. В разд. 2 излагается общий метод построения нетривиальных кривых. В разд. 3 представлен метод построения кривых Гурвица на основе приведения к обобщенным кривым минимального рода по наибольшему показателю степени кривой и оценки сложности вычисления.

1. Основные результаты по кривым Гурвица в конечном поле.

Кривые Гурвица H_n *определяются выражением*

$$X^n Y + Y^n Z + X Z^n = 0, \tag{1}$$

и имеют частные производные $F_X = nX^{n-1}Y + Z^n$, $F_Y = nY^{n-1}Z + X^n$, $F_Z = nZ^{n-1}X + Y^n$.

Несингулярность кривых Гурвица над F_q определяется условиями [7]:

1) n и q должны быть взаимно простыми;

2) $\gcd(n^2 - n + 1, q) = 1$.

Для вычисления рода кривой Гурвица H_n используем выражение для рода кривой $X^a + X^b Y^c + Y^d = 0$ (замечание 4.3 в [8])

$$g \le 1 + \frac{1}{2}\left\{ \left| ac + bd - ad \right| - \gcd(a-b, c) - \gcd(b, c-d) - \gcd(a, d) \right\} \tag{2}$$

Если характеристика поля F_q не делит $\gcd(a-b,c),\gcd(b,c-d),\gcd(a,d)$ и $ac+bd-ad$, тогда справедливо равенство. В силу соотношения (2) для кривой Гурвица H_n выражение для рода имеет вид

$$g=(n^2-n)/2. \tag{3}$$

Существует обобщение кривых Гурвица $H_{n,l}$ вида [9]

$$X^nY^l+Y^nZ^l+X^lZ^n=0, \tag{4}$$

где $n\ge l\ge 2$, $\Delta(n,l)=n^2-nl+l^2\ge 2$ и частные производные $F_X=nX^{n-1}Y^l+lX^{l-1}Z^n$, $F_Y=nY^{n-1}Z^l+lX^nY^{l-1}$, $F_Z=nZ^{n-1}X^l+lY^nZ^{l-1}$.

Несингулярность кривых Гурвица $H_{n,l}$ над F_q определяется условием [9]:

$$\gcd(\Delta(n,l),char(F_q))=1. \tag{5}$$

Род кривой $H_{n,l}$, как следует из (2) и отмечается в [10],

$$g=(n^2-nl+l^2+2-3\gcd(n,l))/2=(\Delta(n,l)+2-3\gcd(n,l))/2. \tag{6}$$

Кривые с числом точек $N\ne q+2$ называются нетривиальными.

Следующая теорема определяет существование нетривиальных кривых Гурвица $H_{n,l}$, для случая $\gcd(n^2-nl+l^2,q-1)=d$ и $\gcd(n,l)=1$.

Теорема 2. [4] Пусть задано конечное поле F_q и $n,l>0$, $\gcd(n,l)=1$. Нетривиальная кривая Гурвица $H_{n,l}$ $X^nY^l+Y^nZ^l+X^lZ^n=0$ существует, если $\gcd(n^2-nl+l^2,q-1)$ содержит делители $d_i^e>3$ такие, что $d_i\equiv 1\bmod 6$, а также делитель равный 3, где $e\ge 1$.

2. Построение нетривиальных кривых Гурвица

Построение нетривиальных кривых Гурвица H_n по делителям порядка поля F_q определяется теоремой 3.

Теорема 3 [4]. Пусть задано конечное поле F_q. Делители порядка поля $q-1$ есть числа $p_1,p_2,...,p_k$ и $p_i\equiv 1\bmod 6$, для $\forall i$ кроме, может быть одного делителя равного 3. Степень n нетривиальной кривой Гурвица $X^nY+Y^nZ+XZ^n=0$ определяется выражением

$$n=n_1P_1+n_2P_2+...+n_kP_k \bmod p_1\cdot p_2\cdot...\cdot p_k, \tag{7}$$

$$P_i=b_i\prod_{\substack{s=1\\ s\ne i}}^{k}p_s\equiv 1(\bmod p_i), \tag{8}$$

где $n_1,n_2,...,n_k$ – образующие элементы мультипликативных подгрупп 6-го и 2-го порядков по модулям $p_1,p_2,...,p_k$, а b_i – целые числа.

Действие теоремы представлено в примере 1.

Пример 1. Пусть задано конечное поле F_q, $q=2^{32}-3713$. Среди делителей порядка поля $q-1$ есть числа $p_1=13$, $p_2=43$, $p_3=7$. Построить по делителям p_1,p_2,p_3 нетривиальную кривую Гурвица H_n.

Решение. Делители p_1,p_2,p_3 определяют мультипликативные подгруппы 6-го порядка, так как $p_i\equiv 1\bmod 6$, $i=\overline{1,3}$. Каждая мультипликативная подгруппа определяется двумя обра-

зующими элементам (по вычислениям из формулы Эйлера). Просто показать, что образующие элементы для подгрупп по модулям p_i, $i=\overline{1,3}$ принимают значения:

- $n_1 = 4$ и 10 для подгруппы по модулю $p_1 = 13$;
- $n_2 = 7$ и 37 для подгруппы по модулю $p_2 = 43$;
- $n_3 = 3$ и 5 для подгруппы по модулю $p_3 = 7$.

Вычисления по формуле (8) дадут следующие значения параметров P_1, P_2, P_3:

- $P_1 = b_1 p_2 p_3 = b_1 43 \cdot 7 = 7 \cdot 301 = 2107 \equiv 1 \bmod 13$;
- $P_2 = b_2 p_1 p_3 = b_2 13 \cdot 7 = 26 \cdot 91 = 2366 \equiv 1 \bmod 43$;
- $P_3 = b_3 p_1 p_2 = b_3 13 \cdot 43 = 6 \cdot 559 = 3354 \equiv 1 \bmod 7$.

Вычисления по формуле (7) по модулю $P_1 \cdot P_2 \cdot P_3 = 13 \cdot 43 \cdot 7 = 3913$ приводят к кривым Гурвица следующего вида:

- $X^{3748}Y + Y^{3748}Z + XZ^{3748} = 0$, для $n_1 = 4$, $n_2 = 7$, $n_3 = 3$;
- $X^{2630}Y + Y^{2630}Z + XZ^{2630} = 0$, для $n_1 = 4$, $n_2 = 7$, $n_3 = 5$;
- $X^{738}Y + Y^{738}Z + XZ^{738} = 0$, для $n_1 = 10$, $n_2 = 7$, $n_3 = 3$;
- $X^{381}Y + Y^{381}Z + XZ^{381} = 0$, для $n_1 = 4$, $n_2 = 37$, $n_3 = 3$;
- $X^{1284}Y + Y^{1284}Z + XZ^{1284} = 0$, для $n_1 = 10$, $n_2 = 37$, $n_3 = 3$;
- $X^{3533}Y + Y^{3533}Z + XZ^{3533} = 0$, для $n_1 = 10$, $n_2 = 7$, $n_3 = 5$;
- $X^{3176}Y + Y^{3176}Z + XZ^{3176} = 0$, для $n_1 = 4$, $n_2 = 37$, $n_3 = 5$;
- $X^{166}Y + Y^{166}Z + XZ^{166} = 0$, для $n_1 = 10$, $n_2 = 37$, $n_3 = 5$.

Замечание 1.

1. Применение теоремы 3 приводит к кривым Гурвица H_n с разными значениям показателя степени n, в зависимости от выбора образующих элементов мультипликативных подгрупп 6-го порядка по модулям $p_1, p_2, ..., p_k$. При этом кривые H_n при различных показателях степени имеют одинаковое число точек (см. утверждение 1 [5]) и разные значения рода (см. выражение (6)).

2. Все кривые из примера 1 являются избыточными по роду, так как параметр $\Delta(n, l=1) = n^2 - n + 1$ имеет не только заданный набор делителей $p_1 = 13$, $p_2 = 43$, $p_3 = 7$. Разложения $\Delta(n,1)$ по делителям имеет вид:

- $\Delta(3748,1) = 3748^2 - 3748 + 1 = 14043757 = 7 \cdot 13 \cdot 37 \cdot 43 \cdot 97$;
- $\Delta(2630,1) = 2630^2 - 2630 + 1 = 6914271 = 3 \cdot 7 \cdot 13 \cdot 19 \cdot 31 \cdot 43$;
- $\Delta(738,1) = 738^2 - 738 + 1 = 543907 = 7 \cdot 13 \cdot 43 \cdot 139$;
- $\Delta(381,1) = 381^2 - 381 + 1 = 144781 = 7 \cdot 13 \cdot 37 \cdot 43$;
- $\Delta(1284,1) = 1284^2 - 1284 + 1 = 1647373 = 7 \cdot 13 \cdot 37 \cdot 43 \cdot 421$;
- $\Delta(3533,1) = 3533^2 - 3533 + 1 = 12478557 = 3 \cdot 7 \cdot 13 \cdot 43 \cdot 1063$;
- $\Delta(3176,1) = 3176^2 - 3176 + 1 = 10083801 = 3 \cdot 7 \cdot 13 \cdot 43 \cdot 859$;
- $\Delta(166,1) = 166^2 - 166 + 1 = 27391 = 7^2 \cdot 13 \cdot 43$.

3. Приведение к кривым наименьшего рода реализуется через обобщенные кривые Гурвица $H_{n,l}$. Существование обобщенных кривых с наименьшим значением параметра $\Delta(n,l) = p_1 \cdot p_2 \cdot p_3$ определяется теоремой 4 [5].

Теорема 4. Пусть задано конечное поле F_q. Делители порядка поля $q-1$ есть числа $p_1, p_2, ..., p_k$ и $p_i \equiv 1 \bmod 6$, для $\forall i$ кроме, может быть, одного делителя равного 3. Тогда существует обобщенная кривая Гурвица $H_{n,l}$ $X^nY^l + Y^nZ^l + X^lZ^n = 0$, такая что $\gcd\left(n^2 - nl + l^2, (q-1)\right) = p_1 p_2 ... p_k$.

Применение теоремы дает следующие нетривиальные кривые:

- $X^{71}Y^{47} + Y^{71}Z^{47} + X^{47}Z^{71} = 0$, для $n_1 = 4$, $n_2 = 7$, $n_3 = 3$;
- $X^{64}Y^{3} + Y^{64}Z^{3} + X^{3}Z^{64} = 0$, для $n_1 = 4$, $n_2 = 7$, $n_3 = 5$;
- $X^{69}Y^{16} + Y^{69}Z^{16} + X^{16}Z^{69} = 0$, для $n_1 = 10$, $n_2 = 7$, $n_3 = 3$;
- $X^{72}Y^{31} + Y^{72}Z^{31} + X^{31}Z^{72} = 0$, для $n_1 = 4$, $n_2 = 37$, $n_3 = 3$;
- $X^{64}Y^{61} + Y^{64}Z^{61} + X^{61}Z^{64} = 0$, для $n_1 = 10$, $n_2 = 37$, $n_3 = 3$;
- $X^{72}Y^{41} + Y^{72}Z^{41} + X^{41}Z^{72} = 0$, для $n_1 = 10$, $n_2 = 7$, $n_3 = 5$;
- $X^{69}Y^{53} + Y^{69}Z^{53} + X^{53}Z^{69} = 0$, для $n_1 = 4$, $n_2 = 37$, $n_3 = 5$;
- $X^{71}Y^{24} + Y^{71}Z^{24} + X^{47}Z^{24} = 0$, для $n_1 = 10$, $n_2 = 37$, $n_3 = 5$.

Замечание 2.

1. Все кривые, рассмотренные в примере, имеют наименьшее значение рода, так как имеют одинаковое значение показателя $\Delta(n,l) = 3913$.

2. Вычисление показателей n и l кривых осуществляется методом последовательного перебора значений наименьшего показателя степени l и вычисления второго показателя по модулю $n' \equiv n \cdot l \bmod p_1 p_2 p_3$ с проверкой разложения на делители $\Delta(n', l)$. Алгоритм останавливается, когда выполнится условие $\Delta(n', l) = p_1 \cdot p_2 \cdot p_3$.

Выводы

1. Теорема 3 не дает прямого ответа на вопрос как построить кривые Гурвица H_n по одному делителю порядка поля. Например, для делителя $p_i = 19$ непосредственные вычисления показывают, что не существует обычной кривой Гурвица H_n, так как не существуют решения в целых числах для уравнения $\Delta(n,1) = 19$, но существуют решения для $\Delta(n,l) = 19$. Таким решением является $n = 5$, $l = 2$.

2. Вычислительные затраты на приведение кривых Гурвица H_n к обобщенным кривым $H_{n,l}$ определяются размером делителей порядка конечного поля и являются пропорциональными произведению этих делителей. Для практически важных конструкций кривых над большими полями $\approx 2^{64} \div 2^{128}$ вычисления становятся значительными. Ниже рассматривается метод построения кривых Гурвица на основе приведения к обобщенным кривым минимального рода.

3. Метод построения обобщенных кривых Гурвица

Многообразие нетривиальных кривых Гурвица определяется значениями делителей порядка поля.

Утверждение 1 [5]. Пусть F_q конечное поле и $q - 1 = p_1^{e_1} p_2^{e_2} ... p_k^{e_k}$, $e_i \geq 1$. Нетривиальные кривые Гурвица $H_{n,l}$ принадлежат одному из семейств:

а) $X^nY + Y^nZ + XZ^n = 0$, если $\Delta(n, l = 1) = n^2 - n + 1 = p_i \cdot ... \cdot p_j$;

б) $X^nY^l + Y^nZ^l + X^lZ^n = 0$, если $\Delta(n,l) = n^2 - nl + l^2 = p_i \cdot ... \cdot p_j$;

в) $X^{cn}Y^{cl} + Y^{cn}Z^{cl} + X^{cl}Z^{cn} = 0$, если $\Delta(cn, cl) == c^2 \cdot p_i \cdot ... \cdot p_j$;

г) $X^c Y^c + Y^c Z^c + X^c Z^c = 0$, если $\Delta(c,c) = c^2$,

где делители $p_1,...,p_j$ тождественны 1 по *mod 6* кроме, делителя равного 3, все $c, p_1,...,p_j$ взяты из набора делителей порядка поля $q-1 = p_1^{e_1} p_2^{e_2} ... p_k^{e_k}$ и $\gcd(n,l) > 1$.

Условия эквивалентности кривых Гурвица определяются следующими утверждениями.

Утверждение 2 [5]. В конечном поле F_q обобщенные кривые Гурвица $X^n Y^l + Y^n Z^l + X^l Z^n = 0$ и $X^n Y^{n-l} + Y^n Z^{n-l} + X^{n-l} Z^n = 0$ являются кривыми одного рода и имеют одинаковое число точек.

Утверждение 3. Кривые Гурвица $X^n Y^l + Y^n Z^l + X^l Z^n = 0$ и $X^l Y^n + Y^l Z^n + X^n Z^l = 0$ являются эквивалентными кривыми одного рода и имеют одинаковые точки с точность до перестановки координат.

Утверждение 2 является очевидным.

Замечание 3.

1. Уравнения а) и б) утверждения 1 определяют кривые Гурвица H_n и $H_{n,l}$. Уравнения в) и г) являются производными от кривых а) и б) и определяются по делителям порядка конечного поля.

2. Утверждения 2 и 3 определяют: можно построить кривую $X^l Y^n + Y^l Z^n + X^n Z^l = 0$ и затем построить классические кривые Гурвица $X^n Y^l + Y^n Z^l + X^l Z^n = 0$ и $X^n Y^{n-l} + Y^n Z^{n-l} + X^{n-l} Z^n = 0$.

Следующие леммы 1 и 2 определяют для заданного значения $\Delta(n,l)$ пределы изменения значения показателя n.

Лемма 1. Параметр $\Delta(n,l)$ лежит в диапазоне

$$n^2 - n^2/4 \le \Delta(n,l) \le n^2 - n + 1. \qquad (9)$$

Действительно минимальное значение $\Delta(n,l)$ определяется в точке $l = n/2$, так как $\left.\frac{\partial \Delta(n,l)}{\partial l}\right|_{l=n/2} = 0$ и справедливо $\Delta(n, l < n) < \Delta(n, l = 1)$.

Лемма 2. Показатель степени кривой Гурвица $H_{n,l}$

$$\sqrt{\Delta(n,l)} < n < 2\sqrt{\Delta(n,l)}/\sqrt{3} \qquad (10)$$

Левая часть соотношения (10) следует из выражения для $\Delta(n,l) \le n^2 - nl + l^2$ и леммы 1. Правая часть определяется выражением для минимального значения $\Delta(n, l = n/2) = 3n^2/4$. ◊

Утверждения 1 – 3 и леммы 1, 2 определяют усовершенствованный метод построения обобщенных кривых Гурвица по заданному набору делителей порядка поля. Основными шагами являются следующие.

1. Фиксируем конечное поле F_q, разложение порядка поля $q-1$ на сомножители, в общем случае, со степенями $q-1 = p_1^{e_1} p_2^{e_2} ... p_k^{e_k}$, $e_i \ge 1$ и набор делителей $p_1,...,p_j$ которые по модулю 6 тождественны единице и, если существует, сомножитель равный 3.

2. Фиксируем делители из набора $p_1,...,p_j$, для которых вычисляем искомое значение параметра $\Delta' = p_1 \cdot ... \cdot p_j$ и формируем набор образующих элементов мультипликативных подгрупп 6-го и 2-го порядков.

3. В соответствии с выражениями (7) и (8) вычисляем набор кривых $X^n Y + Y^n Z + XZ^n$ число которых определяется комбинацией значений образующих элементов мультиплика тивных подгрупп 6-го и 2-го порядков.

4. Для каждой кривой H_n из полученного набора вычисляем $\Delta(n,1)$ и сравниваем с Δ'. Если $\Delta(n,1)=\Delta'$, тогда кривая является без избыточной по роду. Для избыточных кривых $\Delta(n,1)>\Delta'$, переходим к п.5 и 6 – построению обобщенных кривых Гурвица.

5. В соответствии с утверждением 3 уравнения $X^nY+Y^nZ+XZ^n$ из набора кривых п.3 преобразуются в уравнения $XY^l+YZ^l+X^lZ=0$, где $l=n$.

6. Перебираем последовательно значение параметра n от $\lceil\sqrt{\Delta'}\rceil$, где $\lceil\bullet\rceil$ округление до наибольшего целого и вычисляем показатель степени $m=l\cdot n \bmod \Delta'$ для обобщенной кривой $H_{n,m}$ и если $\Delta(n,m)=\Delta'$ – искомая кривая построена. Для построения всех обобщенных кривых следует выполнить вычисления для всех обычных кривых, построенных в п.3 и применить утверждение 2.

Замечание 4. Если кривая Гурвица строится по одному делителю порядка поля p_1, для вычислений следует задать еще один делитель p_2, который по модулю 6 тождественный единице или равный 3. Это не противоречит требованию теорем 3 и 4, так как параметр $\Delta(n,l)$ наряду с искомым делителем может содержать и другие делители не обязательно из делителей порядка поля.

Пример 2. Пусть $p_1=19$. Требуется построить кривую Гурвица $H_{n,l}$ минимального рода.

Зададим дополнительный делитель $p_2=3$. Значение $\Delta'=p_1$.

Образующие элементы для подгрупп по модулям p_i, $i=1,2$ принимают значения:

- $n_1=8$ и 12 для подгруппы по модулю $p_1=19$;
- $n_2=2$ для подгруппы по модулю $p_2=3$;

Вычисления по формуле (8) дадут следующие значения параметров P_1,P_2:

- $P_1=b_1p_2=b_13=13\cdot3=39\equiv1 \bmod 19$;
- $P_2=b_2p_1=b_219=1\cdot19=19\equiv1 \bmod 3$.

Для образующих $n_1=8$, $n_2=2$ по формуле (7) получим $n=n_1P_1+n_2P_2 (\bmod p_1p_2)=8\cdot39+2\cdot19=8 \bmod 57$ и кривую Гурвица

$$X^8Y+Y^8Z+XZ^8=0. \qquad (11)$$

Вычисления для образующих $n_1=12$ и $n_2=2$ приводят к $n=n_1P_1+n_2P_2 (\bmod p_1p_2)=12\cdot39+2\cdot19=50 \bmod 57$ и

$$X^{50}Y+Y^{50}Z+XZ^{50}=0. \qquad (12)$$

Кривые (11), (12) являются избыточными по роду. Для приведения к кривым минимального рода используем метод построения обобщенных кривых. По пункту 5 получим кривые:

- $XY^8+YZ^8+X^8Z=0$;
- $XY^{50}+YZ^{50}+X^{50}Z=0$.

Вычисление обобщенных кривых начинается с нижней границы для $n=\lceil\sqrt{\Delta'}\rceil=5$. Для $XY^8+YZ^8+X^8Z=0$ получим $m=l\cdot n \bmod \Delta'=8\cdot5 \bmod 19=2$ и кривую минимального рода $X^5Y^2+Y^5Z^2+X^2Z^5=0$. Аналогично для кривой $XY^{50}+YZ^{50}+X^{50}Z=0$ получим $m=l\cdot n \bmod \Delta'=50\cdot5 \bmod 19=3$ и обобщенную кривую минимального рода $X^5Y^3+Y^5Z^3+X^3Z^5=0$. Заметим, что обычной кривой Гурвица с $\Delta(n,1)=19$ не существует.

Предложение 1. Метод построения кривой Гурвица $H_{n,l}$ с перебором наибольшего показателя n в 7.46 раза быстрее по сравнению с перебором по наименьшему показателю l.

Действительно, по лемме 2 показатель n искомой кривой $H_{n,l}$ минимального рода лежит в интервале $\sqrt{\Delta(n,l)} < n < 2\sqrt{\Delta(n,l)}/\sqrt{3}$ и максимальное число итераций для п. 6 при вычислении кривой $H_{n,l}$: $T_n = (2-\sqrt{3})\sqrt{\Delta(n,l)}/\sqrt{3}$. В известном методе по теореме 4 [8] вычисления выполнялись по показателю l со значения $l = 2$. Лемма 2, в силу утверждения 3, определяет верхнее значение для показателя степени $l < 2\sqrt{\Delta(n,l)}/\sqrt{3}$. Максимальное число итераций при вычислении кривой $H_{n,l}$ по показателю l равно $T_l = 2\sqrt{\Delta(n,l)}/\sqrt{3}$. Число вычислений при итерациях по l в среднем будет больше числа итераций по n в $T_n / T_l = 2/(2-\sqrt{3}) \approx 7.46$ раз.

Замечание 5. Для кривых $H_{n,l}$ из примера 1 вычисления по предложенному методу в шесть раз быстрее, так как показатель n пробегает значения от $n = \left|\sqrt{\Delta'}\right| = \left|\sqrt{3913}\right| = 63$ до $n = 72$ и заканчивается за 10 шагов, а для итераций по показателю l требуется 60 шагов.

Выводы

1. Предложен метод построения кривых Гурвица на основе приведения к обобщенным кривым минимального рода с последовательным перебором наибольшего показателя кривой Гурвица, который является дальнейшим развитием метода построения кривых Гурвица по делителям порядка поля и отличается от известного введением дополнительного делителя в случае построения кривой по одному делителю порядка поля.

2. Построение кривой Гурвица $H_{n,l}$ по методу приведения к обобщенным кривым минимального рода с перебором наибольшего показателя в среднем в 7.46 раза быстрее по сравнению с перебором по наименьшему показателю степени кривой и расширяет множество кривых, так как определяет построение кривых по одному делителю порядка поля.

Список литературы: 1. *Халимов, Г.З.* Максимальные кривые Гурвица для целей универсального хеширования // Материалы XI Междунар. науч.-практ. конференции «Информационная безопасность». Ч. 3. – Таганрог : Изд-во ТТИ ЮФУ, 2010. – С.144-146. 2. *Халимов, Г.З.* Универсальное хеширование по алгебраическим кривым в простом поле / Г.З.Халимов // Системи управління, навігації та зв'язку / Міністерство промислової політики України, ДП «Центральний науково-дослідний інститут навігації і управління». – Київ. – 2011. – Вип. 1(17). – С.156-161. 3. *Халимов, Г.З.* Универсальное хеширование по максимальным кривым Гурвица / Г.З. Халимов // Прикладная радиоэлектроника. – 2010. – Т.9, № 3. – С.365-370. 4. *Халимов, Г.З.* Условия существования нетривиальных кривых Гурвица / Халимов Г.З. // Системи обробки інформації МО України / Харківський університет Повітряних Сил ім. Івана Кожедуба. – 2010. – Вип. 6(87) – С. 229-233. 5. *Халимов, Г.З.* Условия построения нетривиальных кривых Гурвица / Халимов Г.З. // Системи управління, навігації та зв'язку / Міністерство промислової політики України, ДП «Центральний науково-дослідний інститут навігації і управління». – Київ. – 2010. – Вип 3(15). С..125-129. 6. *Халимов, Г.З.* Кривые Гурвица с большим числом точек в расширенных конечных полях / Г.З.Халимов // Системи управління, навігації та зв'язку. – 2011. – Вип. 2(18). – С.185-189. 7. *Hoholdt, N.* Algebraic geometry codes / N.Hoholdt, J.H. van Lint and R.Pellican // In the Handbook of Coding Theory, (V.S. Pless, W.C. Huffman and R.A. Brualdi Eds.), Elsevier, Amsterdam. – 1998. –V. 1. –P.871-961. 8. *Pellikan, R.* The Klein quartic, the Fano plan and curves representing design / R.Pellikan // In Codes, Curves and Signals: Common Threads in Communications, (A. Vardy Ed.), Kluwer Academy Published, Dordrecht. – 1998. – P.9-20. 9. *Torres, F.* Plan maximal curves / F.Torres // Acta Arithmetic. – 2001. – Vol. 98, No. 2. – P. 165-179. 10. *Beelen, P.* The Newton polygon of plane curves with many rational points. / P.Beelen, R.Pellikan // Designs, Codes and Cryptography. – 2000. – N.21. – P.41–67.

Харьковский национальный университет радиоэлектроники

Поступила в редколлегию 20.09.2012

УДК 621.391:519.2:519.7

А.П. БУБИР, І.Д. ГОРБЕНКО, д-р техн. наук

ОЦІНКА СТІЙКОСТІ НАПРАВЛЕНОГО ШИФРУ NTRU ДО АТАКИ З АДАПТИВНО ПІДІБРАНИМИ ШИФРОТЕКСТАМИ

Вступ

Використання криптосистем з відкритим ключем – основний спосіб захисту різноманітної інформації, що передається за допомогою комп'ютерних мереж. Найбільш широко використовуються такі криптосистеми, RSA яких заснована на складності факторизації великих чисел, схема Діфі – Гелмана, DSA, стійкість яких базується на складності вирішення задачі дискретного логарифмування в полі, сімейство алгоритмів у групі точок еліптичних кривих. Але всі вони мають певні недоліки, основні з яких – це або відносно низька швидкість (наприклад, алгоритми на базі ЕК), або низька стійкість (схема Діфі – Гелмана та інші алгоритми, засновані на проблемі пошуку дискретного логарифма в полі), або ці два недоліки одночасно (RSA). Вирішити проблему низької швидкості шифрування при збереженні високого рівня стійкості покликаний алгоритм NTRU, який було розроблено в середині 1990-х років та вперше представлено на конференції CRYPTO'96. NTRU було запатентовано компанією «NTRU Cryptosystems» (зараз – «Security Innovation») 24 липня 2000 р. В цьому алгоритмі всі операції виконуються у кільці зрізаних поліномів. Криптографічна стійкість алгоритму заснована на складності знаходження короткого вектора в заданій алгебраїчній решітці. В 2008 р. даний алгоритм був включений в стандарт [1], а його модифікована версія була взята за основу стандарту [2]. NTRU виконує операції шифрування, розшифрування, генерації ключів на порядки швидше [3] у порівнянні з іншими асиметричними криптосистемами, однак його криптографічна стійкість до різних класів атак ще недостатньо вивчена.

Опис атаки

В цій роботі була розглянута аналітична атака з адаптивно підібраними шифротекстами. Атака можлива за умови, що криптоаналітик отримає деяке число криптограм e, які являють собою одне й те ж повідомлення m, зашифроване на одному й тому ж відкритому ключі h, але із використанням різних сеансових ключів («засліплюючих поліномів») r_i [2]. Тоді криптоаналітик зможе відновити більшість або всі коефіцієнти полінома r_0, який відповідає криптограмі e_0, і потім обчислити значення зашифрованого повідомлення.

Нижче наведено узагальнений алгоритм атаки:

1) криптоаналітик перехоплює або будь-яким іншим чином отримує набір із C криптограм, які мають вищенаведену властивість;

2) на наступному кроці він обчислює різниці $r_i - r_0, 0 < i < C$ згідно з наступним співвідношенням:

$$(e_i - e_0) * h_q^{-1} = r_i - r_0 (\mathrm{mod}\, q), \tag{1}$$

де e_i – i-та криптограма.

Справедливість виразу (1) витікає з наступного. Якщо e_i визначимо як:

$$e_i = r_i * h + m (\mathrm{mod}\, q), \tag{2}$$

то

$$e_i - e_0 = (r_i * h + m) - (r_0 * h + m) (\mathrm{mod}\, q) = (r_i - r_0) * h (\mathrm{mod}\, q). \tag{3}$$

Помноживши обидві частини рівності (3) на інверсію відкритого ключа h_q^{-1}, отримуємо формулу (1);

3) аналізуючи коефіцієнти отриманих різниць, криптоаналітик проводить процедуру відновлення коефіцієнтів r_0. Якщо r_0 має число одиниць та мінус одиниць, яке дорівнює значенню d_r використаного набору параметрів, то він відновлений правильно, у противному випадку алгоритм сигналізує о про невдалу спробу розкриття й переривається;

4) на останньому кроці криптоаналітик обчислює значення вихідного повідомлення m:

$$m = e_0 - r_0 * h(\mathrm{mod}\, q). \qquad (4)$$

Опишемо більш детально ідею, яка лежить в основі процедури відновлення коефіцієнтів r_i. Якщо поліноми r_j є бінарними (коефіцієнти приймають значення {0, 1}), то кожний коефіцієнт різниці $r_i - r_0$ буде належати множині {-1, 0, 1}. Якщо він дорівнює 0, то даний факт не дасть додаткової інформації про значення відповідного коефіцієнта r_0. (1 – 1 = 0, 0 – 0 = 0). Але різниця, яка дорівнює 1, можлива лише тоді, коли коефіцієнт r_i дорівнює одиниці, а коефіцієнт r_0 дорівнює нулю (1 – 0 = 1). Аналогічно, різниця, яка дорівнює мінус одиниці, означає, що r_i має на даній позиції нуль, а r_0 – одиницю. Кількість одиниць та мінус одиниць у отриманих різницях визначає, чи буде атака успішною. Як тільки буде виявлено, що поліном, який відновлюється, містить d_r одиниць, то всі нерозкриті позиції заповнюються нулями, алгоритм зупиняється. На виході отримуємо сеансовий ключ r_0.

Якщо коефіцієнти r_j приймають значення з множини {-1, 0, 1}, то їх різниця може дорівнювати -2, -1, 0, 1, 2. Однозначно дозволяють визначити коефіцієнти сеансового ключа r_0 тільки значення -2 (-1 – 1) та 2 (1– (-1)). У першому випадку коефіцієнт r_0 дорівнює одиниці, у другому – мінус одиниці. Треба зауважити, що у випадку тернарних коефіцієнтів криптоаналітику необхідно використати більшу кількість криптограм (ймовірність того, що окремо взятий коефіцієнт різниці дозволить відновити коефіцієнт r_0, дорівнює 2/5, в бінарному випадку така ймовірність значно вища й складає 2/3).

Опис інструментального засобу

При програмному моделюванні було використано мову С й відкриту бібліотеку libntru, яка реалізує всі основні математичні операції в кільцях зрізаних поліномів, а також процедури генерації ключової пари, направленого шифруванння за алгоритмом NTRU згідно із чорновою версією стандарту [1].

Вибір libntru у якості основної бібліотеки для реалізації атаки обумовлений тим, що на сьогоднішній день вона є єдиною бібліотекою з відкритим вихідним кодом, яка реалізує алгоритм НШ NTRU. Поточна версія бібліотеки працює тільки під UNIX-подібними операційними системами (GNU/Linux, MacOS X, FreeBSD, Android й т.п.). Це пояснюється використанням у якості джерела псевдовипадкових бітів спеціальних символьних псевдопристроїв /dev/random або /dev/urandom (на вибір користувача), які відсутні в сімействі ОС Windows Компіляція даної бібліотеки під Windows потребує модифікації коду, який реалізує генерацію псевдовипадкових послідовностей. Для роботи деяких функцій libntru потрібна наявність у системі криптографічного пакету OpenSSL.

Практична робота складалася з двох етапів. Перший етап – реалізація алгоритму атаки з адаптивно підібраними шифротекстами на мові С. Ця програма моделює процес взаємодії між валідним користувачем та криптоаналітиком. Зі сторони користувача випадковим чином генерується повідомлення m, виконується його зашифрування по алгоритму NTRU із використанням декількох сеансових ключів r_j. Потім вважається, що набір таких криптограм потрапив у розпорядження криптоаналітика. Програма починає відновлення коефіцієнтів «засліплюючого» полінома r_0 та обчислює згідно із (3.4) повідомлення-кандидат m'. На останньому кроці проводиться порівняння вихідного повідомлення m та результату атаки m' виводиться повідомлення про їх спів падіння чи неспівпадіння.

Другий етап полягав у визначенні залежності між ймовірністю неправильного відновлення відкритого тексту $P_{пом}$ та кількістю криптограм спеціального виду C, які доступні зловмиснику. З цією метою на основі програми, яка реалізує атаку, було створено тест, який

для кожного значення C із заданого діапазону проводить серію спроб здійснення атаки й підраховує кількість невдалих результатів. Потім обчислюється ймовірність помилки, яка дорівнює відношенню кількості невдалих результатів до кількості усіх спроб проведення атаки для заданого C. Виходом програми тестування є набір пар $\{C, P_{пом}\}$.

Запуск програми тестування здійснюється із командного рядка. Перед запуском необхідно явно вказати місце розташування динамічної бібліотеки *libntru.so*, якщо вона не була раніше скопійована в один із стандартних каталогів, що зберігають бібліотеки (/lib, /usr/lib в ОС Linux). Дана процедура виконується шляхом встановлення змінної оточення LD_LIBRARY_PATH:

$ export LD_LIBRARY_PATH="місце_розташування_бібліотеки".

Програма тестування має повноцінний інтерфейс командного рядка, який дозволяє задати всі основні параметри атаки. Формат запуску програми:

$ blindings_test [-N max_power] [-f num_ones] [-m mess_num_ones] [-b] [-C ct_num] [-e num_of_test_iters] [-h].

Більш детальну інформацію щодо призначення опцій програми можна отримати, задавши при її виклику ключ –h.

```
ander@ander-laptop:~/libntru$ export LD_LIBRARY_PATH="."
ander@ander-laptop:~/libntru$ ./blindings_test -N 1087 -f 120 -m 120 -C 7 -e 100
 -b
----- Starting test -----
Number of experiments: 100
Number of fails: 3
Failure probability: 0.030000
----- End of testing -----
ander@ander-laptop:~/libntru$
```

Рис. 1. Приклад запуску тесту (ОС Linux)

Результати роботи програми:

– Number of experiments – кількість проведених експериментів (визначається параметром -e);

– Number of fails – кількість невдалих результатів експериментів;

– Failure probability – ймовірність помилки відновлення вихідного повідомлення.

Результати досліджень

Дослідження проводились з метою визначення мінімальної кількості криптограм, при наявності якої ймовірність неправильного відновлення повідомлення прямує до нуля. Тестування проводилось для наборів параметрів ees401ep1, ees449ep1, ees677ep1, ees1087ep1 із стандарту [2], які є представниками всіх можливих рівнів стійкості (112, 128, 192 та 256 біт відповідно). Ймовірність обчислювалась як для бінарного, так і для тернарного представлень сеансових ключів. Такий вибір методики тестування обумовлений необхідністю визначення загальносистемних параметрів, від яких може залежати ймовірність невдачі, а також тим фактом, що в теорії для відновлення коефіцієнтів, які мають значення {0, 1} за їх різницями знадобиться менша кількість вибраних криптограм (див. п. 3.1).

За отриманими результатами було побудовано графіки залежності ймовірності помилкового відновлення повідомлення $P_{пом}$ від кількості шифротекстів C для бінарного й тернарного випадків (рис. 2, 3).

Експериментальні дані підтвердили теоретичне припущення про те, що для успішного відновлення бінарних коефіцієнтів сеансового ключа криптоаналітику необхідно одержати меншу кількість криптограм спеціального вигляду. Наприклад, для набору параметрів ees1087ep1 (рівень стійкості 256 біт) ймовірність помилки відновлення бінарних коефіцієнтів наближається до нуля вже при кількості криптограм, рівній 10, тоді як для тернарних поліномів цей показник наближається до 90.

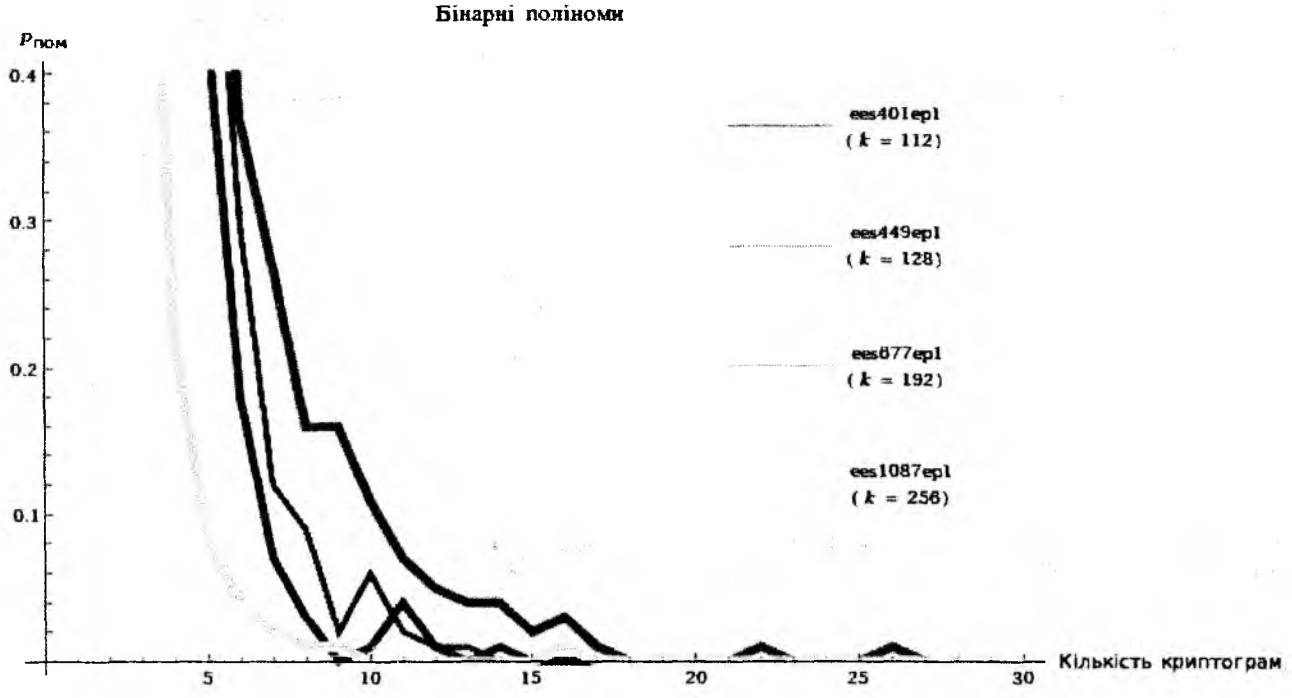


Рис. 2. Залежність ймовірності невдалого результату атаки від кількості шифротекстів (бінарні поліноми)

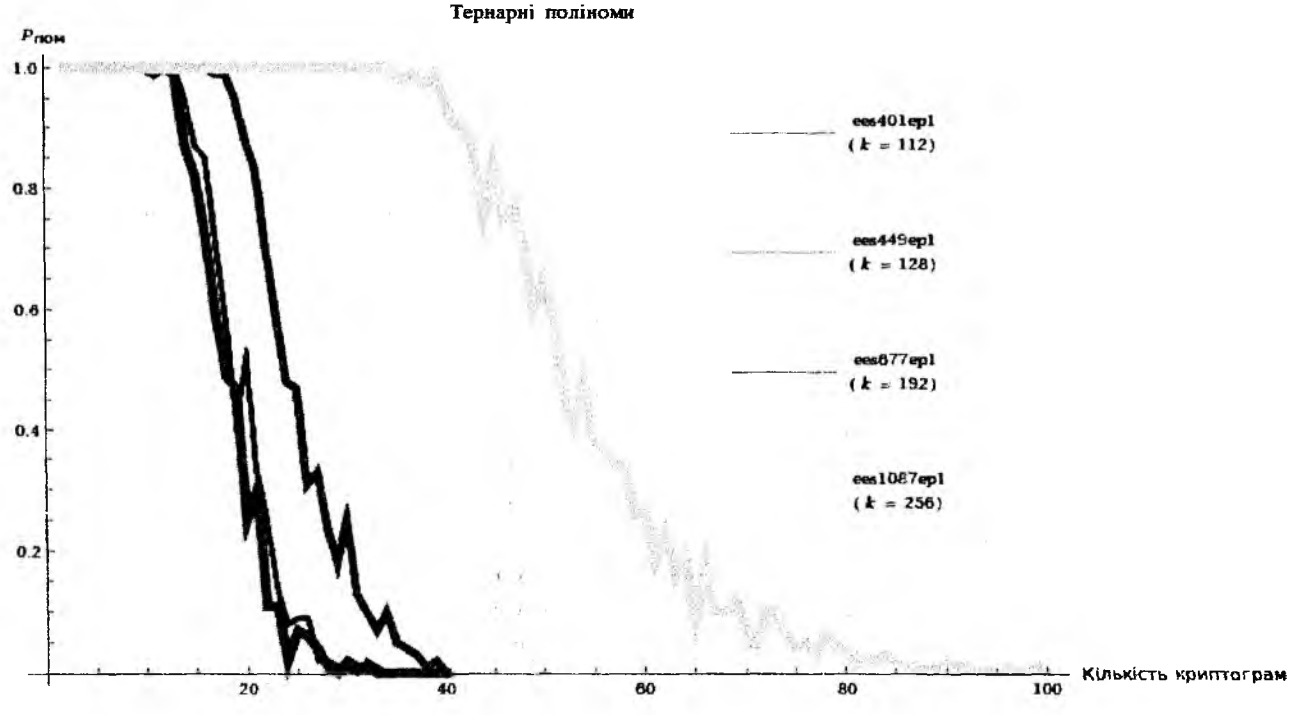


Рис. 3. Залежність ймовірності невдалого результату атаки від кількості шифротекстів (тернарні поліноми)

Було також виявлено, що у випадку бінарних коефіцієнтів мінімальна кількість шифротекстів, при якій $P_{пом} \approx 0$, залежить тільки від кількості одиниць d_f в сеансовому ключі r. Максимального значення вона досягає для набору параметрів ees667ep1, де $d_f = 157$, мінімального – для наборів ees401ep1 и ees1087ep1 (113 и 120 одиниць відповідно). Параметр N,

який визначає рівень стійкості, для бінарних коефіцієнтів практично не впливає на ймовірність успішної реалізації атаки, а обумовлює лише незначне збільшення часу її виконання із зростанням N (через обчислення різниць векторів, які мають більшу кількість коефіцієнтів).

Для тернарних коефіцієнтів, на відміну від двійкових, чітко відстежується зв'язок між рівнем стійкості параметрів та величиною N (див. рис. 3). Збільшення N поряд із незначним зростанням часу атаки призводить до пропорційного збільшення кількості криптограм, необхідних для їх успішної реалізації.

Рекомендації із захисту від атаки з адаптивно підібраними шифротекстами

Єдиний спосіб убезпечити поточну версію алгоритму НІІ NTRU від такої атаки, який гарантовано не вплине на стійкість до інших видів атак, – не пересилати те ж саме повідомлення, зашифроване на одному й тому ж відкритому ключі, більш ніж один раз. Для стандарту ANSI X9.98 ця вимога повинна виконуватись «апріорі», тому що вважається, що алгоритм NTRU буде використовуватись виключно для безпечного обміну ключами симетричного шифрування, які є випадковими. В цьому випадку ймовірність повтору повідомлення m повністю визначається властивостями джерела ключових даних для симетричних криптоалгоритмів.

Ще один метод захисту від можливості проведення саме цієї атаки – використання одного й того ж «засліплюючого» полінома r для пари {повідомлення, відкритий ключ}. Згідно із стандартом [2] r є результуючим значенням геш-функції

$$r(x) = H(m(x) \| b), \tag{5}$$

де b – сіль – бітова послідовність, яка повинна формуватися випадковим чином.

Для забезпечення константного значення r необхідно, щоб сіль b стала не випадковою, а детермінованою. В такому випадку будь-яке повідомлення m буде при повторному шифруванні відображатися в ту ж саму криптограму e. Ця умова несумісна із семантичною стійкістю та стійкістю від атаки із вибраними шифротекстами (зловмисник не повинен мати можливість взнати будь-яку інформацію про відкритий текст, якщо він має в розпорядженні лише криптограму). Ми прийшли до протиріччя: захист від атаки з підібраними шифротекстами таким методом зробить шифр вразливим до цілого ряду інших атак.

Висновки

Стійкість до атаки з адаптивно підібраними шифротекстами можлива тільки тоді, коли для кожного повідомлення m генерується унікальне, єдине значення полінома r. У зв'язку з тим, що r є результатом використання геш-функції до конкатенації повідомлень m із сіллю b, ймовірність того, що для двох різних пар (m, b) буде згенеровано одне й те ж значення r, повністю визначається ймовірністю колізії значень геш-функції. Повторне шифрування повідомлення на одному й тому ж відкритому ключі неприпустиме. В іншому випадку шифр є абсолютно нестійким до атаки із адаптивно підібраними шифротекстами. Тестування програмної моделі атаки показало, що в середньому достатньо перехопити 30-40 таких криптограм, щоб, не знаючи особистий ключ, відновити початкове повідомлення.

Таким чином, направлений шифр NTRU, як і всі інші класи асиметричних алгоритмів, підпорядковується фундаментальному правилу побудови криптосистем, яке забороняє використання НІІ для безпосереднього шифрування повідомлень.

Список літератури: 1. *IEEE* 1363.1 IEEE Standard Specification for Public-Key Cryptographic Techniques Based on Hard Problems over Lattices [Електронний ресурс]. – 2008. 2. *American* National Standard for Financial Services ANSI X9.98-2010 Lattice Based Polynomial Public Key Establishment Algorithm for the Financial Services Industry [Електронний ресурс]. – 2010. 3. *Jens Hermans, Frederik Vercauteren, Bart Preneel.* Speed records for NTRU [Электронный ресурс]. – Режим доступа: www / URL: https://www.securityinnovation.com/uploads/ntru_speed_benchmark_research.pdf. – 2010.

Харківський національний університет радіоелектроніки

Надійшла до редколегії 15.09.2012

УДК 621.391:519.2:519.7

Х.А. БУГАЄНКО, І.Д. ГОРБЕНКО, д-р техн. наук

АНАЛІЗ ДВОХ МЕТОДІВ АВТЕНТИФІКАЦІЇ ОСОБИ ДЛЯ ЗАСТОСУВАННЯ В ЕЛЕКТРОННОМУ БІОМЕТРИЧНОМУ ПАСПОРТІ

Вступ

В проекті закону України " Про державний демографічний реєстр " в якості основних способів запропоновано два способи автентифікації особи на основі підпису і біометричних ознак геометрії обличчя. Перший зі способів запропонований давно і набув широкого застосування в дактилоскопії та перший випадок відноситься до 1900 р. Перше застосування автоматизованого розпізнавання осіб по геометрії обличчя відноситься до 60-х років минулого століття. На наш погляд і нині актуальним є аналіз та порівняння цих методів з метою вибору більш ефективного. Зрозуміло, що вживання ефективного вимагає введення показників ефективності і це також актуальна задача. Більш актуальною являється є задача ефективності двох способів, коли вони застосовуються одночасно.

Метою статті є оцінка і порівняльний аналіз методів біометричної автентифікації, що базуються на використанні підписів та геометрії обличчя, а також оцінка ефективності у випадках, коли для автентифікації застосовуються два методи одночасно.

Основні галузі застосування біометричних даних

Інтеграція Україні до Європи зачіпає не лише економічну і політичну, але і соціальну сферу життя наших громадян. Європейський союз визначив набір протоколів для реалізації EAC (доступу до критичних даних (біометричних даних) в своїх електронних паспортах (ЕП, ePassport)). В Україні і в країнах ЄС ведеться робота з впровадження біометричних паспортів.

Основними російськими стандартами з біометричної автентифікації є ГОСТ Р ISO/IEC 19794-6–2006, ГОСТ Р ISO/IEC 19794-5–2006, ГОСТ Р ISO/IEC 19794-2–2005, ГОСТ Р ISO/IEC 19794-4–2006, ГОСТ Р ISO/IEC 19785-1–2008, ГОСТ Р ISO/IEC 19794-7–2009, ГОСТ Р ISO/IEC 19795-2–2006.

Біометричний паспорт – це документ, що дає право на виїзд за межі країни і в'їзд до іноземних країн. Основною відмінністю електронних документів від існуючих паперових аналогів є те, що в них може бути внесений біометричний набір характеристик, замінити які важче ніж надрукований набір даних, і відповідно набагато важче видавати себе за власника паспорта. Біометричні дані є особливо критичною інформацією, доступ до якої мають отримувати виключно ті системи перевірки, які можуть підтвердити свої повноваження на дані дії. Основними джерелами з питань біометричного паспорту є різноманітні публікації організації цивільної авіації ІСАО.

Автентифікація на основі біометричних даних – це засіб автоматичного розпізнавання особи на базі унікальних фізичних або поведінкових параметрів людини. Автентифікація виконується за допомогою порівняння отриманих біометричних характеристик і шаблонів (еталонів), що зберігаються у базі даних, в залежності від характеристик, які при цьому використовуються.

Біометричні системи поділяються на статичні та динамічні. До методів першої групи відносяться такі види автентифікації: за відбитком пальця, за формою долоні, за розташуванням вен на лицьовій стороні долоні, по сітківці ока, по райдужній оболонці ока, за геометрією обличчя, за термограмою обличчя. за ДНК, інші методи. До динамічних методів відносять автентифікацію за рукописним почерком (динамічний підпис), за клавіатурним почерком, за голосом, інші методи. В даний час активно використовуються такі біометричні ознаки, як відбитки пальців, геометрія обличчя, особливості голосу, райдужна оболонка ока, рукописний (динамічний) підпис.

Автентифікація осіб за біометричними ознаками геометрії обличчя

Обличчя – це передня частина голови, на ньому знаходяться: очі, брови, ніс, рот, щоки, лоб і підборіддя. Обличчя дуже чітко висловлює (міміка) людські емоції, такі як біль, радість, розчарування, втому, неспокій, злість. Обличчя є найбільш індивідуальною частиною тіла людини, тому саме по його цілісному сприйняттю відбувається ідентифікація однієї особи іншою [1]. Автоматизоване розпізнавання осіб – відносно нове поняття. Початок розвитку бере з 60-х років минулого століття, перша напівавтоматична система для розпізнавання осіб прагнула, щоб адміністратор визначив місцезнаходження особливостей обличчя (таких як очі, вуха, ніс і рот) на фотографіях перш, ніж було розраховано відстані і відношення до загального орієнтиру (еталонного зображенням на даний момент), які були тоді, в порівняні з сьогоденням, довідковими даними.

Фотографія та цифровий опис особи заносяться в базу даних, з якої згодом порівнюється розпізнаване особа. На основі набору біометричних даних і їх обробки формується якийсь електронний зразок як унікальне у своєму роді число, що відноситься до конкретної особистості.

Вибір і виділення ознак грають в розпізнаванні образів ключову роль. Є три категорії ознак: фізичні, структурні і математичні.

Фізичні та структурні ознаки виражаються через форми обличчя (овал особи, геометрія його основних частин), його колір, а також колір волосся і т.д. Найбільш часто використовуваний ознака – яскравість. До фізичних і структурним ознаками можна віднести також координати точок обличчя в місцях, відповідних зміні контрасту (брови, очі, ніс, вуха, рот і овал).

До математичних ознак належать спектри вихідних образів, характеристики, градієнти зміни яскравості та інше, статистичні отримані в результаті математичного перетворення вихідних ознак [Біометрія. Зображення обличчя [2].

Інформаційні знаки обличчя [2]:

– форма обличчя;

– співвідношення частин обличчя між собою (лоб, середня і нижня частини обличчя);

– форма чола, скул і підборіддя;

– симетрія / асиметрія особи;

– форма і розмір вуха, спосіб його прикріплення, форма частин вуха

– форма, величина, розташування очей, рота, носа;

– лінії зморшок.

Методи автентифікації [5]:

1. Метод збирання характеристик

Способи виділення розрізняються залежно від використовуваного способу представлення і способу його редукції (скорочення).

Ознаки яскравості збираються в вектор з допомогою перерахуванням значень яскравості кожного пікселя. Для образу 112 × 92 пікселів виходить матриця порядку 10304. У такому просторі пошук практично неможливий, тому зображення попередньо зменшується до прийнятних розмірів.

Серед математичних ознак найчастіше використовуються спектральні ознаки, отримані в результаті перетворення Фур'є. Розмір початкового простору ознак може скласти до 200 елементів. При цьому для більш високого рівня розпізнавання перетворення Фур'є застосовуються не для усередненого образу, а для кожного окремо.

Аналіз локальних ознак. На образі виділяються координати особи і локальних ознак (наприклад, куточки рота, ніс, очі і т.д.) Координати ознак і відстані між ними дозволяють описати особа за допомогою точок і параметрів, які потім будуть використовуватися при розпізнаванні.

2. Кореляційний метод

Найпростіший з вище перелічених методів. Якщо умови отримання нових образів відповідають умовам отримання еталона (освітлення, пункт спостереження обличчя, нахил, масштаб, фон і т.д.), то кореляція (відповідність) між ними близька до одиниці. Рівень розпізнавання доходить до 96 %.

3. Метод на основі перетворень Карунена – Лоєва (ПКЛ)

Цей метод дозволяє значно скоротити розмір еталона, залишаючи тільки ті ознаки, які мають принципове значення для конкретного образу. Рівень правильного розпізнавання стабільно досягає 80 %, навіть при сильних змінах умов. На рис. 1 приведена структурна схема даного методу.

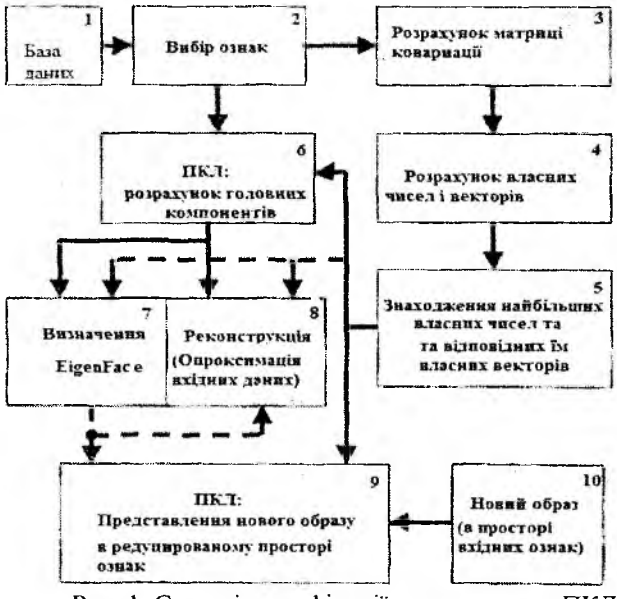


Рис. 1. Схема ідентифікації за допомогою ПКЛ

4. Метод на основі лінійного дискримінантного аналізу (ЛДА)

LDA – статистичний підхід для того, щоб класифікувати зразки по класах, що засновані на навчальних зразках з відомими класами [3]. Даний метод, як і попередній, дозволяє скоротити кількість ознак, при цьому істотно покращує класифікацію образів (розрізнення одного від другого). Це дозволяє збільшити рівень розпізнавання – до 99%, навіть при різних умовах.

Для біометричного методу ідентифікації особи за біометричними ознаками геометрії обличчя розглянемо всі ймовірні переваги та недоліки, які містяться в даному методі (дивитись табл.1).

Таблиця 1

Переваги та недоліки методу

Переваги методу	Недоліки методу
- низька ціна пристрою отримання зображення; - безконтактність; - зручна і швидка обробка даних; - можливість використання самонавчаючихся систем (нейроні мережі).	- складність реалізації системи; - залежність відеообразів від перешкод; - повина забезпечуватись висока якість зображення.

Автентифікація осіб за рукописним підписом

Підпис – один із класичних способів ідентифікації, застосовуваний вже кілька століть в юридичній практиці, банківській справі і торгівлі.

Перші згадки про даний метод датуються приблизно 1900 роком в одній з перших виданих робіт Осборна.

Автоматична перевірка підпису є важливою областю досліджень у зв'язку з соціальними та правовими поняттями та широке використання рукописного підпису як особистий метод перевірки автентичності особи. Даний метод легкий у використані – це основна його перевага. Є два види підробки підпису: випадкова (ненавмисна) підробка підпису та навмисна (з використанням в корисних цілях).

Є два незалежних способу ідентифікації по підпису:

1. Ідентифікація по малюнку підпису на документі.

2. Ідентифікація по динаміці підпису, що вводиться в комп'ютер.

У першому способі порівнюються два зображення. З цим краще справляється людина. У другому способі є дані про коливання пера при відтворенні підпису в тривимірному просторі (X, Y – координати і Z – тиск на папірець/планшет). З цим може справитися тільки комп'ютер. На рис. 2 наведено динаміку написання підпису.

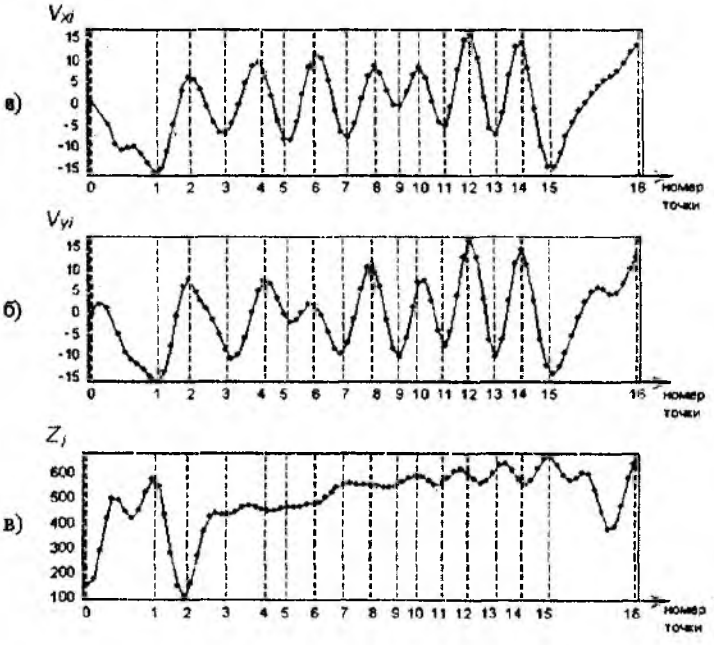


Рис. 2. Криві, що характеризують динаміку написання підпису

Методи перевірки підпису можуть бути класифіковані відповідно з вхідним відомостями про підпис, поділяються на два класи [5]:

1. Он-лайн.

2. Офф-лайн.

Он-лайн метод відноситься до використання функції часу динамічного процесу підписання (наприклад, тиск на ручку залежно від часу), яка отримується за допомогою пристроїв, таких як сенсорні екрани або планшети.

Off-лайн метод відноситься до використання статичного зображення підпису.

Системи перевірки рукописного підпису [5]:

1. Збір даних і первинна обробка. Збір даних можливий при наявності відповідних пристроїв – таблетки переводу підпису в цифрову форму, сенсорні екрани або планшети. Спеціальні засоби для визначення тиску ручки на папір під час підписання. Кроки попередньої обробки підпису перед виділенням ознак – фільтрування шумів (наприклад, з гаусівськими вікнами [4]) і передискретизація.

1.1. Реєстрація, запис. Залежно від стратегії відповідно реєстрація може бути розділена на два класи: заснована на пошуку та заснована на моделі.

Реєстрація заснована на пошуку – зображення вхідного підпису (еталонного) по черзі порівнюється з базою підписів.

Реєстрація заснована на моделі – використовується набір (база даних) підписів для оцінки статистичної моделі, яка описує поведінку особливості виконання підпису особою. Велика проблема, пов'язана зі стадією реєстрації, є зміна часу підпису.

1.2. Обчислення подібності

1.2.1. Попереднє вирівнювання. Стадії відповідності взагалі передує попереднє вирівнювання між вхідним підписом та зареєстрованим шаблоном / моделлю (еталонним підписом) дивись рисунок 3.

1.2.2. Узгодження

1.2.3. Оцінка нормалізації. Оцінки одержаної відповідності відбувається шляхом порівняння вхідного підпису з шаблоном, як правило, нормований на загальний діапазон, такий як [0, 1], перш ніж порівнюючи їх з порогом ухвалення рішення, використовуються різні функції відображення. Цей крок розрахунку нормалізації використовується при об'єднанні різних обчислювачів в біометричному підході.

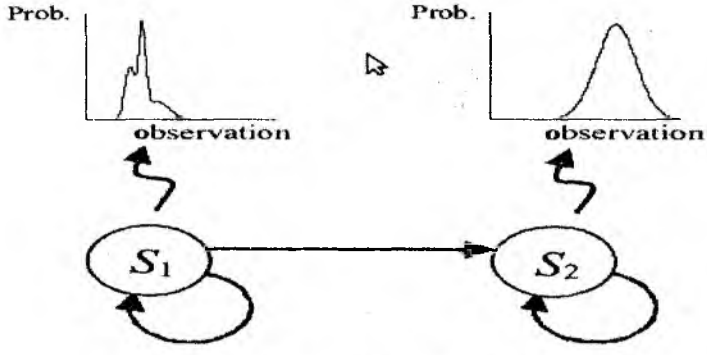


Рис. 3. Ілюстрація вирівнювання

2. Скелетне уявлення рукописної кривої. В основі методу аналізу динамічних кривих лежить ідея виділення характерних або екстремальних точок [1]. При цьому ключове значення мають так звані «вертикальні» екстремуми, що характеризують зміна напрямку руху пера у вертикальній площині. На рис. 4 наводиться приклад «скелетного» уявлення рукописного підпису, отриманий шляхом з'єднання екстремальних точок кривої.

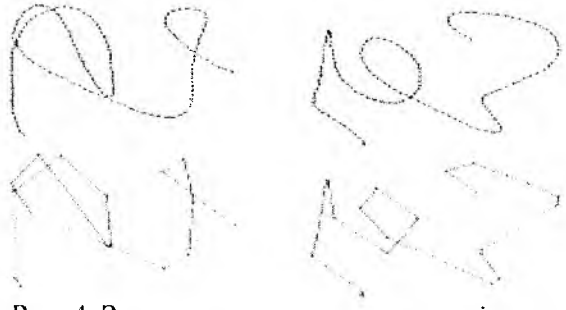


Рис. 4. Зразок скелетного уявлення підпису

Заміна вихідної кривої ламаною лінією досить точно описує поведінку траєкторії дозволяє формалізувати процедуру зіставлення двох кривих.

3. Загальна схема підпрограми пошуку еталонного підпису. З метою дотримання оптимального балансу між продуктивністю підсистеми пошуку і якістю її роботи був реалізований ієрархічний,багаторівневий підхід.

Першим кроком підпрограми пошуку був відбір можливих кандидатів для наступного аналізу. При цьому брався до уваги наступний набір характеристик:

– пропорції (ширина і висота) підпису;
– координати центру мас підпису;
– кількість сегментів;
– кількість петель;
– максимальна довжина штриха в вертикальному та горизонтальному напрямках;
– кількість вертикальних екстремумів;
– відношення числа ліво орієнтованих штрихів до право орієнтованим.

На наступному кроці алгоритму знайдені кандидати передавалися на вхід процедурі пошук. Алгоритм пошуку відбирає три кандидати, які доставляють мінімальні штрафи за невідповідність. На третьому кроці знайдені три гіпотези перевіряються на структурну відповідність з пред'явленим зразком. І тільки на заключному етапі порівнювані рукописні криві представляються за допомогою сплайнів Безьє, і таким чином робиться спроба оцінити ступінь схожості еталона і зразка по їх формі. У табл. 2 проаналізовано та наведено можливі недоліки та переваги методу автентифікації особи за рукописним підписом.

Таблиця 2

Переваги та недоліки методу

Переваги методу	Недоліки методу
- невисока вартість; - підпис- біометричний параметр людини; - процес верифікації дуже швидкий; - для зберігання шаблонів необхідно мало місця; - на результат не впливає мова, якою говорить особа; - стиснення зображення не впливає на результат.	- високий рівень помилок 1і 2 роду; - час реєстрації близько двох хвилин; - підпис не надійний спосіб автентифікації та захисту; - висока ймовірність підробки підпису.

Висновки

Автентифікацію з використанням рукописного (динамічного) підпису не можна використовувати всюди – зокрема, цей метод не підходить для обмеження доступу в приміщення або для доступу в комп'ютерні мережі. Однак у деяких областях, наприклад в банківській сфері, паспорті громадянина України, а також там, де відбувається оформлення важливих документів, перевірка правильності підпису може стати найбільш ефективним, а головне – необтяжливим і непомітним способом.

Імовірність помилки 1-го роду становить близько 0,01 – це цілком прийнятно. Однак для помилок 2-го роду – це багато. Серед переваг можна відзначити невисоку вартість і відносну звичність для людини. До недоліків відносяться високий рівень помилок 1 і 2 роду, тривалий час реєстрації користувача (більше двох хвилин).

Метод розпізнавання з використання біометричних ознак геометрії обличчя – це єдиний біометричний спосіб ідентифікації особи з точки зору багатоцільового застосування. На відміну від інших біометричних методів, застосовних для контролю доступу або порівняння в базі даних, технологія розпізнавання образу дозволяє детектувати (знаходити) особу для подальшого порівняння з базою даних. Завдяки вбудованим інфрачервоним випромінювачам комп'ютер легко розпізнає каучукову маску як муляж, що імітує особа. Об'єкт, що проходить автентифікацію, повинен мати людську шкіру, природну міміку і бути "живим", в іншому випадку спрацьовує звуковий сигнал попередження.

Помилкова відмова доступу (помилкове заперечення): від 10 до 20 на 100 звернень, складає приблизно 10 – 20 % [6].

Помилковий доступ (помилкове визнання): від 100 до 1000 на 100000 звернень до системи, це відповідає 0,1 – 1 % [6].

У табл. 3 наводяться характерні оцінки помилок першого та другого роду для автентифікації особи по геометрії обличчя.

Таблиця 3

Результати методу

Назва методу	FAR	FRR
Марківська модель	2,9 %	31 %
Метод вейвлет Габора	2,9 %	30 %

Як і аналіз по особистому рукописному підпису, визначення ідентичності по фотографії в паспорті громадянина України відносять до найбільш доступного і визнаного методу розпізнавання особистості. Застосування двох методів для автентифікації особи за біометричними ознаками геометрії обличчя та за рукописним (динамічним) підписом в паспорті громадянина України будуть набагато ефективніші, якщо застосовувати їх одночасно. Так як ймовірність підробки біометричних даних зменшується в декілька разів в порівняні з окремим - застосуванням кожного з методів. Одночасно зменшаться й помилки як першого, так і другого роду, що призведе до підвищення ефективності і надійності розглянутих двох методів.

Список літератури: 1. *Болл, Р.М.* Основні біометричні параметри // Керівництво по біометрії. – Москва, 2007. – Г. 3. – 47-65 с. 2. *Електронний ресурс* // Режим доступу: http://wiki.oszone.net/index.php / Біометрія. Зображення обличчя. 3. *Bolme, D, Beveridge, R., Teixeira, M. and Draper B.* The CSU Face Identification Evaluation System: Its Purpose, Features and Structure // International Conference on Vision Systems, Graz, Austria, April 1-3, 2003. (Springer-Verlag). – 304-311 p. 4. *Baron, R. and Plamondon, R.* Acceleration measurement with an instrumented pen for signature verification and handwriting analysis. IEEE Trans. on Instrum. Measurement, 38(6):1132–1138, 1989. 5. *Anil, K. Jain.* Handbook of Biometrics. Springer Science+Business Media, 2008. 6. *Biometrics catalog* // Режим доступу: http://www.biometricscatalog.org/.

Харківський національний університет радіоелектроніки

Надійшла до редколегії 07.09.2012

УДК 621.391:519.2:519.7

Ю.І.ГОРБЕНКО, О.С.ТОЦЬКИЙ, В.А.ПОНОМАР

АНАЛІЗ МЕТОДІВ ЗНЕОСОБЛЕННЯ ПЕРСОНАЛЬНИХ ДАНИХ

Вступ

Зараз, в час розвитку електронного документообігу, коли майже вся інформація зберігається та обробляється в електронному вигляді одним з найважливіших питань становиться питання захисту цієї інформації. Особливо це стосується персональних даних.

Надання послуг конфіденційності є однією з головних вимог до органів які зберігають персональні дані. Майже всі вони зберігаються в базах персональних даних і за Законом України «Про захист персональних даних» її власник повинен забезпечити безпеку персональних даних, що в ній знаходяться.

Головною загрозою для бази даних є виток інформації, що в ній зберігається, тому постає задача захисту даних від витоку, тобто забезпечення її конфіденційності. Це можливо за рахунок використання різних програмних, апаратних чи програмно-апаратних засобів.

Але іноді для баз даних різних установ виникають випадки коли потрібно передати персональні дані на обробку, не надаючи доступ до самої інформації. Одним способом вирішення цієї задачі є шифрування інформації перед тим як відіслати її на обробку. Але що робити коли потрібно відіслати базу даних на обробку без доступу до інформації, але зберігши її внутрішні зв'язки та різні статистичні властивості. В даному випадку шифрування не може використовуватися. І саме в таких випадках виконується знеособлення персональних даних.

Мета та задачі знеособлення персональних даних

При роботі з персональними даними трапляються випадки коли потрібно ці дані знеособити. Для того щоб зрозуміти принципи та методику знеособлення, треба проаналізувати мету та задачі знеособлення:

1) Метою знеособлення персональних даних, що містяться у базах даних (БД), є можливість передачі такої БД для виконання тестування із навантаженням за межами України [1].

2) Метод знеособлення персональних даних має бути погоджений із Державною службою України з питань захисту персональних даних [2].

3) Метод знеособлення персональних даних повинен надавати можливість знеособлення персональних даних для БД розміром 150 Тб.

4) Метод знеособлення персональних даних може включати можливість виконання переіндексації даних БД.

5) Метод знеособлення персональних даних повинен залишати структуру БД, розміри даних, внутрішні логічні зв'язки БД.

Методи реалізації знеособлення персональних даних

Загалом знеособлення персональних даних окремих БД виконується за рахунок підстановки замість значення у полі БД частини гами (псевдовипадкової послідовності – ПВП), що генерується за алфавітом притаманним для даного типу даних.

Гама управління в залежності від вимог може генеруватися двома методами. Основний метод відповідає ДСТУ 4145-2002 [3]. Він полягає у тому, що генератор за одне звернення до нього видає випадковий рядок довжини t=0. Як криптографічне перетворення в генераторі застосовується алгоритм криптографічного перетворення згідно з ГОСТ 28147 у режимі простої заміни [4]. Таблиця заміни і особистий ключ цього перетворення повинні відповідати ГОСТ ГОСТ 28147-89. Умови отримання й використання особистого ключа повинні унеможливлювати доступ до нього або його частини, модифікацію, підміну або знищення. Особис-

тий ключ криптографічного перетворення згідно з ГОСТ 28147, що використовується в генераторі випадкових послідовностей, не можна використовувати для іншої мети. Сам алгоритм полягає в наступному:

1) Задається початковий стан генератора – 64 біти.

2) Обчислюється $I = E_k(D)$ (де E_k – шифрування згідно ГОСТ 28147).

3) $x = {}_{Ek}(I \oplus s)$ (де $\oplus$ – сума по модулю 2 бітових рядків)

4) $s = E_k(x \oplus I)$

Необхідний для гами біт – молодший біт бітового рядка *x*. Таким чином генерується необхідна кількість біт гами підстановки.

Другий, перспективний метод, дозволяє генерувати гаму з необхідними властивостями, його відмінність в тому, що алфавіт гами може бути довільний [5]. Він зводиться полягає у тому, що у загальному випадку детермінований генератор ПВП, що функціонує згідно трьох модульного перетворення в розширенні поля Галуа $F(p^n)$ може бути реалізований на основі рекуретного обчислення згідно

$$b_i = \left((\theta_j)^i \left(mod(f(X), p, n), (f_1(X), p_1, n_1), (f_m(X), \vec{m})\right)\right), \tag{1}$$

$$b_i = \left((\theta_j)^{k_0+i} \left(mod(f(X), p, n), (f_1(X), p_1, n_1), (f_m(X), \vec{m})\right)\right). \tag{2}$$

Для знеособлення персональних даних застосовується метод генерування ПВП на основі багатомодульного перетворення елементів тільки в простому полі Галуа GF (p), тобто при n=1.

1) Ввести або генерувати загальносистемні параметри p та θ побудування елементів поля Галуа, де p – просте число, а θ – первісний (породжуючий просте поле Галуа) елемент поля, а також просте число p_1.

2) Ввести параметр генератора (ключ), k=1 ÷ p-1.

3) Обчислити початкове значення генератора a_0:

$$a_0 = \theta^k (mod\ p) = R_p(\theta^k). \tag{3}$$

4) Обчислити z елементів a_i ПВП:

$$a_i = a_{i-1}\theta (mod\ p) = R_p(a_0\theta^i)_{i=1,2,3,\ldots,z} \tag{4}$$

5) Обчислити елемент b_i послідовності {B}, що генерується:

$$b_i = a_i (mod\ p_1) = R_{p_1}(a_i) = R_{p_1}(R_p(a_i\theta^i)). \tag{5}$$

Таким чином одержуємо z блоків ПВП.

Для заміни необхідно використання гами підстановки – ПВП із необхідними властивостями необоротності, нерозрізнюваності, непередбачуваності та заданим (гарантованим) періодом повторення. При цьому статистичні властивості символів гами підстановки повинні співпадати зі статистичними властивостями символів вихідної мови, а метод генерування гами повинен опиратись на діючі в Україні стандарти криптографічного захисту інформації.

В знеособлює мій базі підстановки повинні виконуватися для таких алфавітів:

1) цифри (0-9), основа алфавіту m=10;

2) український алфавіт, заголовні букви (А-Я), основа алфавіту m=33;

3) український алфавіт, заголовні та прописні букви (А-Я, а-я), основа алфавіту m=66;

4) український алфавіт, заголовні та прописні букви, цифри (А-Я, а-я, 0-9), основа алфавіту m=76;

5) український алфавіт, заголовні букви, англійський алфавіт, заголовні букви, цифри (А-Я, A-Z, 0-9), основа алфавіту m=69;

6) англійський алфавіт, заголовні та прописні букви, символи (A-Z, a-z, \), основа алфавіту m=53;

7) англійський алфавіт, заголовні та прописні букви, цифри (A-Z, a-z, 0-9), символи (,_,.,), основа алфавіту m=64.

Метод генерування ПВП c_i з алфавітом m зводиться до виконання таких кроків:

1. Якщо m =10, то генеруємо байти, тобто числа в діапазоні 0 -255 діляться на 10 інтервалів і число із відповідного інтервалу заміняється на відповідну цифру.

2. Якщо m =33 (український алфавіт), то згідно табл. 1 частот появи 33 символів української мови формуємо випадкові послідовності символів згідно таблиці і заміни випадкового байта на символ алфавіту з урахуванням частоти його появи.

Таблиця 1

Прог./0,134	о/0,082	н/0,070	а/0,070	и/0,056	т/0,051	в/0,046
е/0,043	р/0,038	і/0,037	с/0,036	к/0,036	м/0,033	д/0,028
л/0,028	у/0,027	п/0,026	я/0,021	з/0,019	ь/0,015	г/0,013
ч/0,011	б/0,010	х/0,010	ц/0,009	ю/0,009	ж/0,008	й/0,007
ї/0,006	є/0,006	ф/0,005	ш/0,005	щ/0,003	ґ/0,000	

3. Якщо m =66 (український алфавіт, заголовні та прописні букви), то згідно табл. 1 частоти появи символів української мови формуємо випадкові послідовності символів згідно сформованої таблиці і заміни випадкового байта на символ алфавіту з урахуванням частоти його появи. Частота появи заголовної букви береться в залежності від даних поля БД.

4. Якщо m =76 (український алфавіт, заголовні та прописні букви, цифри), то згідно табл. 1 частот появи символів української мови формуємо випадкові послідовності для символів української мови згідно таблиці. Частота появи цифр чи заголовних букв береться в залежності від даних поля БД, а вірогідність появи відповідної цифри вважати рівно ймовірною відносно інших.

5. Якщо m =69 (український алфавіт, заголовні букви, англійський алфавіт, заголовні букви, цифри 0-9). Основа алфавіту m =69, то частоти появи 33 заголовних букв української мови визначаємо згідно табл.1, частоти появи 26 заголовних букв англійської мови згідно табл.. 2, появу цифр вважати рівно ймовірною. Частота появи букви відповідної мови чи цифри залежить від даних поля БД.

Таблиця 2

Прог./0,137	e/0.127	t/0.097	i/0.075	a/0.073	o/0.068	n/0.067
s/0.067	r/0.064	h/0.049	c/0.045	l/0.040	d/0.031	p/0.030
y/0.027	u/0.024	m/0.024	f/0.021	b/0.017	g/0.016	w/0.013
v/0.008	k/0.008	x/0.005	q/0.002	z/0.001	j/0.001	

6. Якщо m =53 (англійський алфавіт, заголовні та прописні букви, та символ '/'), то згідно з табл. 2 частот появи символів англійської мови формуємо випадкові послідовності для символів української мови згідно з таблицею. Частота появи заголовних букв береться в залежності від даних поля БД, а символ '/' в процесі знеособлення будемо залишати незмінним.

7. Якщо m =64 (англійський алфавіт, заголовні та прописні букви, цифри, та символи '_' та '.'), то згідно з табл. 2 частот появи символів англійської мови формуємо випадкові послідовності для символів української мови згідно таблиці. Частота появи цифр чи заголовних букв береться в залежності від даних поля БД, вірогідність появи відповідної цифри вважати

рівно ймовірною відносно інших, а символи '_' та '.' в процесі знеособлення будемо залишати незмінним.

Таким чином, у залежності від значення *m*, генерування послідовності заміни здійснюється у такій послідовності:

- генерується двійкова ПВП необхідної довжини;
- двійкова послідовність перетворюється в послідовність необхідного числа бітових блоків, довжина яких задається в залежності від вимог до знеособлення та властивостей способу генерування ПВП;
- у відповідності з частотами появи букв та цифр бітові блоки замінюються на необхідні букви, цифри, символи відповідних алфавітів.

Оскільки зв'язки БД виконанні через ідентифікатори, які не несуть в собі якусь інформацію про данні, то вони знеособленню не підлягають і через це зв'язки БД не руйнуються.

Висновки

На сьогодні реалізація знеособлення персональних даних є одним за важливіших питань в даній сфері. Найважливішими вимогами до знеособлення персональних даних є вимога зберігання зв'язків БД, що виконується майже завжди, оскільки ідентифікатори не несуть якоїсь інформації про данні і тому не виникає потреба в їх знеособлені. Друга вимога полягає в тому, щоб знеособлена БД мала ті ж статистичні властивості, що й її оригінал і саме цю вимогу виконати найскладніше. Для зрозуміння цієї проблеми наведемо приклад роботи методу з найкращими статистичними даним – він виробляє алфавіт не для поля, а для символу при цьому використовується лише три алфавіти: англійський, український, цифри. Заміна символу в ньому виконується на символ того ж алфавіту, крім того заголовна буква замінюється на заголовну, прописна – на прописну. Різні інші символи залишаються незмінні. Результат роботи цього методу наведено в табл. 3.

Таблиця 3

№	Тип даних	Приклад даних	Приклад результату
1	Паспортні дані		
а	ПІБ	Іванов Владислав Андрійович	Еунг оуЯмтсохваняАарезидитп
б	Серія та № паспорта	МН987430	РО259692
в	Адреса реєстрації	61172,м.Харків, вул. Зубарєва 61, кв.15	56112,ио.Рбиачо,пктт. Е в нгосл74,ову.35
г	Дата народження	21.02.1992	83.03.8334
2	Номер телефону (можливе збереження префіксу у початковому вигляді)	0983728814	8510353088
3	Імена облікових записів користувачів системи	biks_iv_2.narod.ua	n__fa84a0.x_u_r.om
4	E-mail адреса абонентів та користувачів системи	Vladlv@gmail.com	afcr_l@bninp.ocd

Як результат при знеособленні дійсної БД статистичні властивості результату максимально наближені до оригіналу. Великим мінусом цього методу є його швидкодія, оскільки перевіряється кожний символ, то при знеособленні БД об'ємом в декілька Тб її обробка може зайняти декілька місяців, коли метод який був наведений в попередніх пунктах буде працювати вдвічі швидше оскільки алфавіт вибирається не на символ, а на поле БД (а то й колонку даних).

Є ще способи підвищення швидкодії знеособлення. По-перше, при розбивці блоку гами на інтервали (щоб виконати підстановку символу згідно частоти його появи) зробити масив (таблицю) підстановки і далі не перевіряти значення блоку гами, в який інтервал він потрапляє, а використовувати його як індекс таблиці підстановки. Це збільшить швидкодію, але потребує значно більший об'єм пам'яті. що використовується при знеособленні. По-друге,

перед знеособленням провести відповідну підготовку БД. Вивести дані з спільним алфавітом в окремий розділ і забрати їх назад після знеособлення – це збільшить швидкодію бо потрібно буде вибирати алфавіт лише на початку розділу, а не при кожному переході до наступного поля. Недоліками цього способу є те, що його потрібно застосовувати в дійсно великих БД (1 Тб і більше) інакше він навпаки зменшить швидкодію, крім цього реалізація цього способу на великих даних досить складна.

Наведені методи – це методи знеособлення персональних даних шляхом шифрування його змісту. Вони залишають всі властивості БД незмінними, але вилучають інформацію, що надає змогу ідентифікувати об'єкти персональних даних, що відповідає вимогам закону України про захист персональних даних.

Список літератури: 1. *Закон України* про захист персональних даних №2297 від 01.06.2010. 2. *Закон України* "Про захист інформації в інформаційно-телекомунікаційних системах" (із змінами). 3. *ДСТУ 4145-2002* «Інформаційні технології. Криптографічний захист інформації. Цифровий підпис, що ґрунтується на еліптичних кривих. Формування та перевіряння». 4. *ДСТУ ГОСТ 28147:2009* «Системы обработки информации. Защита криптографиеская. Алгоритм криптографического преобразования». 5. *Горбенко, І.Д., Горбенко, Ю.І.* Прикладна криптологія. Теорія. Практика. Застосування : монографія. – Харків : Форт, 2012.

Харківський національний
університет радіоелектроніки

Надійшла до редколегії 12.09.2012

УДК 681.3.06

А.В. ПОТИЙ, д-р техн. наук, Д.Ю. ПИЛИПЕНКО

МОДЕЛЬ ИНСТИТУЦИОНАЛЬНОГО УПРАВЛЕНИЯ ДЕЯТЕЛЬНОСТЬЮ ПО ЗАЩИТЕ ИНФОРМАЦИИ

Введение

Сегодня можно с уверенностью утверждать, что понимание проблем управления процессами защиты информации (ЗИ) подвергается качественному изменению. Работающие в данном направлении исследователи отмечают важность организационных (нетехнических аспектов) защиты информации наравне с техническими [1 – 3]. С точки зрения управления и оценки, организационные аспекты ЗИ (человеческий фактор, культура информационной безопасности) представляют собой значительную сложность. Одновременно с этим присутствует острая потребность в разработке моделей управления деятельностью по ЗИ, которые учитывают организационные аспекты ЗИ.

Исследование деятельности по ЗИ осуществляется в рамках новых теоретических моделей и подходов. К таким моделям и подходам относятся: процессный подход, игровая модель ЗИ, программно-целевой подход, сценарный подход, рациональная модель ЗИ, ситуационный подход, институциональное управление. Модель институционального управления деятельностью по ЗИ в рамках исследования организационных аспектов представляет значительный интерес.

В работе приводятся результаты онтологического моделирования института информационной безопасности (ИИБ) и предметной области культуры информационной безопасности (КИБ). Предлагается модель институционального управления деятельностью по ЗИ Формулируются выводы и рекомендации относительно применения институционального подхода в деятельности по ЗИ.

1. Предметная область института и культуры информационной безопасности

Ключевой функцией ИИБ является согласование целей деловой деятельности организации с задачами безопасности и профессиональными обязанностями персонала. На рис. 1 приведена онтологическая модель терминологического ядра предметной области ИИБ, раскрывающая взаимосвязь между основными понятиями (компонентами).

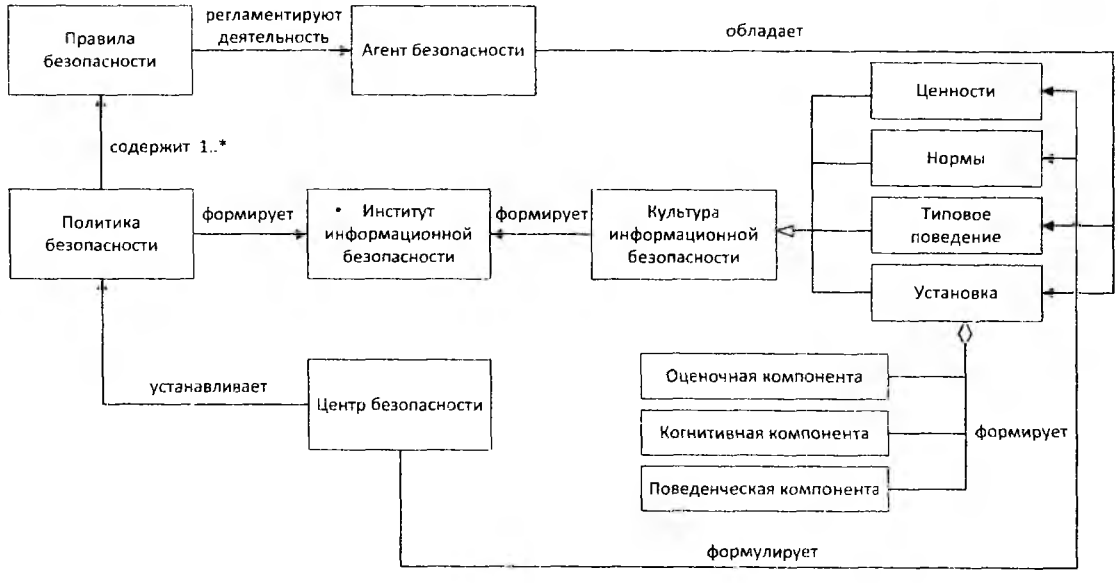


Рис. 1. Онтологическая модель предметной области института и культуры информационной безопасности

Анализ приведенной модели позволяет предложить следующее определение понятия "институт информационной безопасности":

Определение 1.1. *Институт информационной безопасности это упорядоченная система ценностей, норм и правил безопасности в контексте защиты информации, формирование которого направлено на согласование деятельности агента безопасности с целями и задачами, установленными центром безопасности.*

Как видно из онтологической модели (рис. 1), на формирование ИИБ оказывают влияние два субъекта деятельности по ЗИ: центр безопасности и агент безопасности. Центр безопасности может быть представлен руководством организации, начальником службы безопасности или другими управляющими субъектами, чьи интересы должны быть защищены. Агент безопасности представлен сотрудником организации или другим лицом, которое не принимает непосредственного участия в процессе формирования управляющих воздействий.

Определение 1.2. *Центр безопасности – это субъект, который осуществляет управление деятельностью по защите информации, проводит целенаправленное формирование политики безопасности и культуры информационной безопасности с конечной целью сформировать институт информационной безопасности.*

Определение 1.3. *Агент безопасности – это управляемый субъект деятельности по защите информации, который оказывает влияние на формирование культуры информационной безопасности посредством своего поведения и установок по отношению к требованиям безопасности.*

Культура информационной безопасности является не только существенным фактором, который оказывает влияние на общую защищенность организации, но и частью более сложной структуры – института информационной безопасности. В рамках модели институционального управления предлагается рассматривать КИБ как инструмент управления деятельностью по ЗИ. Сущность институционального управления заключается в формировании норм и правил безопасности. Нормы и правила могут быть выражены в явной и неявной форме. Тогда политику безопасности можно отнести к группе явных форм, а КИБ отнести к неявной форме выражения норм и правил. Следует подчеркнуть, что в организационном аспекте политика безопасности представляет собой механизм принуждения, в то время как КИБ представляет собой механизм побуждения. Предлагается следующее определение КИБ:

Определение 1.4. *Культура информационной безопасности – это набор норм, ценностей, установок и стандартов, которые формируют допустимое поведение в контексте деятельности по защите информации.*

2. Модель институционального управления безопасностью информации

Модель институционального управления безопасностью информации представлена на рис. 2. Данная модель состоит из четырех блоков: мотивационный аспект, модель центра безопасности, модель агента безопасности, модель оценки уровня КИБ. Субъектами деятельности по ЗИ в данной модели выступают центр безопасности (управляющий субъект) и агент безопасности (управляемый субъект).

2.1. Мотивационный аспект деятельности по ЗИ.

Не вызывает сомнений, что в основе любой человеческой деятельности лежит определенная потребность. Наличие потребности формирует у субъекта деятельности мотив и цель. В контексте ЗИ потребность проявляется в том, что субъект деятельности по ЗИ ощущает необходимость в обеспечении такого состояния информации, при котором не существует угроз информационным активам организации или же их уровень приемлем. Цель отражает желаемый результат, достижение которого способно удовлетворить потребностям субъекта деятельности по ЗИ. Мотив выступает побудителем к действию, являясь тем, ради чего осуществляется деятельность по ЗИ.

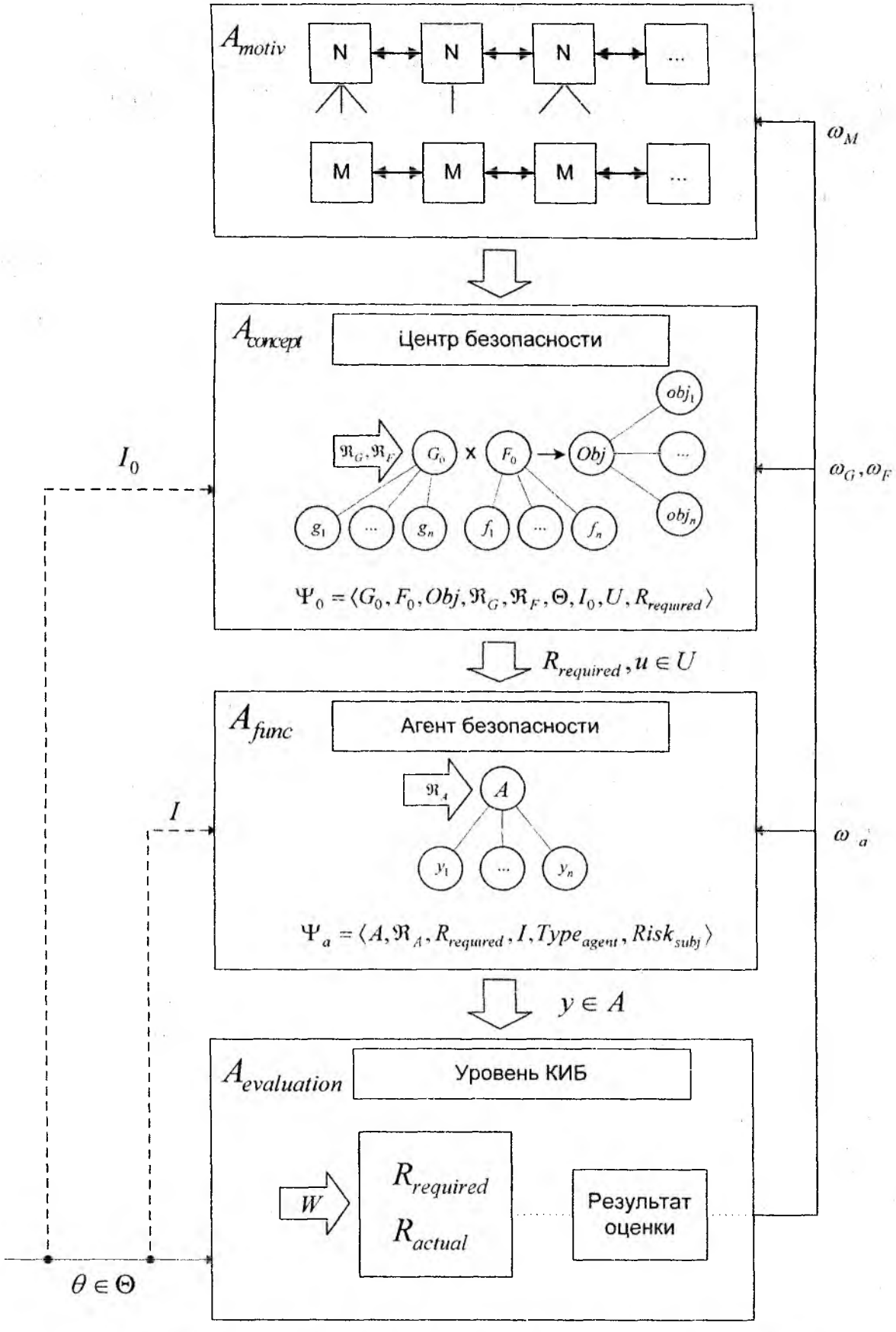


Рис. 2. Модель институционального управления деятельностью по ЗИ

Центр обладает рядом потребностей и, как следствие, совокупностью мотивов. Как видно из блока A_{motiv} (рис. 2), количество связей между потребностью и мотивом может варьироваться. Это зависит как от специфики организации, в которой находится субъект деятельности по ЗИ, так и от индивидуальных особенностей его как личности. Наличие множества потребностей N и множества мотивов M приводит к формированию концепции обеспечения безопасности информации (блок $A_{concept}$).

2.2. Модель центра безопасности

Деятельность центра безопасности заключается в реализации некоторого управляющего воздействия $u(\cdot) \in U$ по отношению к агенту. Как видно из рис. 2, выбор стратегии $u(\cdot)$ из множества допустимых стратегий U является результирующим этапом процесса принятия решения центром. Принято считать, что центр не производит результат деятельности, не опосредованный агентом. Роль центра заключается в выборе типа управления и его реализации, поэтому результатом деятельности центра принято считать результат деятельности агента [4]. Выбор управляющего воздействия описывается моделью принятия решений центром безопасности Ψ_0.

Помимо управляющего воздействия $u(\cdot) \in U$, модель принятия решений центра безопасности включает в себя: множество целей ЗИ G_0, множество факторов F_0, функции предпочтений $\Re_G, \Re_F$ на множестве целей и множестве факторов соответственно, множество задач безопасности $Obj = \{obj_1, obj_2, ..., obj_n\}$, множество обстановок Θ, информацию о конкретной обстановке на момент принятия решения I_0, требуемые результаты деятельность по ЗИ. Модель принятия решений центром можно описать следующим кортежем:

$$\Psi_0 = \left\langle G_0, F_0, Obj, \Re_G, \Re_F, \Theta, I_0, U, R_{required} \right\rangle \quad (1)$$

На данном этапе центр формулирует главные цели защиты информации G_0 и в случае необходимости осуществляет их декомпозицию на требуемое число подцелей $g_i \in G$. Расставляя приоритеты между целями, центр использует свои предпочтения на множестве целей $\Re_G$. Пространство целей ЗИ является в достаточной мере гибким, поэтому декомпозиция целей может осуществляться как на этапе проектирования, так и непосредственно в ходе осуществления деятельности по ЗИ. Достижение целей ЗИ g_i происходит под влиянием множества факторов F_0. Аналогично множеству целей, множество факторов также подвергается декомпозиции. Центр осуществляет ситуационный анализ и выделяет наиболее существенные факторы, опираясь на свои предпочтения на множестве факторов $\Re_F$. Соотношение целей $g_i \in G$ и факторов $f_i \in F$ формирует множество задач защиты информации $Obj = \{obj_1, obj_2, ..., obj_n\}$. Решение задач ЗИ является механизмом осуществления деятельности по ЗИ посредством выполнения конкретных действий (набора операций).

Таким образом, опираясь на множество задач безопасности $Obj = \{obj_1, obj_2, ..., obj_n\}$, центр безопасности может сформулировать требуемый результат $R_{required}$.

Содержательно в данной модели требуемый результат $R_{required}$ будем понимать как уровень культуры информационной безопасности, которая формируется на основе поведения агентов, являющихся результатом управляющих воздействий центра безопасности.

Отклонение фактического результата R_{actual} от требуемого $R_{required}$ может быть обусловлено влиянием обстановки на действия агента: реализацией внешних или внутренних угроз, системными сбоями, действиями других субъектов системы ЗИ и т.д. Данная зависи-

мость выражается функцией $r = W_I(\cdot) = (y, \theta)$, где $r \in R_{actual}$ фактический результат деятельности агента, $y \in A$ действие агента, и $\theta \in \Theta$ частная обстановка на момент действия. Информация, доступная центру о частной обстановке, определяется переменной I_0.

Таким образом, результатом работы модели принятия решений Ψ_0 является некоторое управляющее воздействие $u(\cdot)$. В общем виде задача центра заключается в выборе оптимального воздействия $u(\cdot)$ из множества допустимых альтернатив $u(\cdot) \in U$. В зависимости от своих предпочтений, ограничений на ресурсы и возможностей центр выбирает наиболее рациональное управляющее воздействие на момент принятия решения.

2.3. Модель принятия решений агентом безопасности.

Как уже упоминалось, агент безопасности играет роль управляемого субъекта деятельности по ЗИ. Для того чтобы лучше понять принцип выбора агентом тех или иных действий, необходимо описать модель принятия решений агентом (блок A_{func}). В общем виде данная модель описывается кортежем вида:

$$\Psi_a = \left\langle A, \Re_A, R_{required}, w(\cdot), I, Type_{agent}, Risk_{subj} \right\rangle. \quad (2)$$

Пусть агент безопасности обладает предпочтениями $\Re_A$ на множестве действий A и способен выбирать действие $y \in A$. В зависимости от своего типа $Type_{agent}$ и степени субъективного восприятия риска $Risk_{subj}$, агент склоняется к выбору тех или иных действий. Содержательно тип агента $Type_{agent}$ можно интерпретировать как его манеру взаимодействия с центром: благожелательность, нейтральность или противодействие. Переменная $Risk_{subj}$ выступает мерой субъективного восприятия риска агентом лично. Например, агент с низким уровнем восприятия риска может проигнорировать принятые в организации правила (политику) и нормы безопасности в том случае, когда негативные последствия для него лично не имеют существенного значения.

При выборе действия $y \in A$ агент руководствуется собственными предпочтениями на множестве действий A. Функция $W_I(\cdot)$ устанавливает зависимость между выбором действия $y \in A$, обстановкой $\theta \in \Theta$ и результатом деятельности по ЗИ $r \in R_{actual}$. Тогда выбор действия агента определяется правилом индивидуального рационального выбора:

$$P^{W_I}(Type_{agent}, Risk_{subj}, I, A) \subseteq A \quad (3)$$

Как видно из выражения (3), на выбор агентом безопасности действия $y \in A$ влияет тип агента $Type_{agent}$, мера его субъективного восприятия риска $Risk_{subj}$, информация об обстановке на момент принятия решения I, множество допустимых действий A. Пользуясь данным правилом, из всего множества допустимых действий A агент формирует подмножество наиболее предпочтительных с его точки зрения действий $\tilde{A} \in A$.

Описания поведения агента безопасности будем осуществлять в рамках следующих гипотез [4]: *гипотезу рационального поведения агента* и *гипотезу детерминизма*. Первая гипотеза заключается в том, что агент на основе всей имеющейся у него информации выберет действие, ведущее к наиболее предпочтительным для него результатам. Вторая гипотеза означает стремление агента устранить существующую неопределенность и принимать решения в условиях полной информированности.

2.3.1. Гипотеза рационального поведения агента безопасности.

Согласно гипотезе рационального поведения агент выбирает решение из множества альтернатив, на которых достигается максимум его функции полезности:

$$f(y) = v(w(y,\theta)) \tag{4}$$

Следует отметить, что между действием агента $y \in A$ и результатом $r \in R_{actual}$ не существует *однозначной* связи. Это объясняется тем, что помимо поведения агента, на формирование КИБ также влияет обстановка, которая не всегда связана с действиями агента. Принимая решение, агент стремится прогнозировать результаты своей деятельности с учетом информации, которая отражена в переменной I. Например, на соблюдение требований безопасности работника бухгалтерии повлияет его знание о том, проводится ли систематический контроль его действий. Знание о мониторинге электронной почты может предотвратить нарушение требований безопасности другим агентом. Подобная информированность неизменно отразится на решении, принимаемом агентом безопасности.

С целью осуществить наиболее эффективное действие, максимизирующее функцию полезности, агент стремится устранить неопределенность. Неопределенность, в зависимости от направленности на объекты или субъекты, может быть объективной или субъективной соответственно. Объективная неопределенность касается параметров обстановки, т.е. условий, которые агент учитывает при принятии решения. Данная неопределенность устраняется увеличением информированности агента с помощью параметра I. Например, доведение до сведения сотрудника правил работы с электронной почтой. Субъективная неопределенность представляет собой неполную информированность агента о принципах поведения других субъектов. Уменьшить данную неопределенность способен центр при помощи *рефлексивного управления*, т.е. предоставляя агентам информацию о параметрах других участников системы ЗИ. Например, сотрудники информируются о нарушениях безопасности, произошедших по вине их коллег.

2.3.2. Гипотеза детерминизма.

В случае детерминированного изменения результата деятельности информация об обстановке является несущественной для агента, поскольку в данном случае результат зависит только от действия агента. Иными словами, каждому действию агента соответствует только один результат деятельности $f(y) = v(w(y))$. Правило индивидуального рационального выбора в данном случае будет выглядеть следующим образом:

$$P^{W_I} = (Type_{agent}, Risk_{subj}, A) = Arg \max_{y \in A} f(y) \tag{5}$$

Можно сказать, что в таких обстоятельствах существующая неопределенность устранена, и агент безопасности принимает решение в условиях полной информированности. В этом случае влияние внешней природы на формирование КИБ, и оказать влияние на ее текущее состояние может только агент своим поведением.

Гипотеза детерминизма на практике практически нереализуема, поэтому центр безопасности принимает все управляющие воздействия в предположении о рациональности поведения агента безопасности.

2.4. Модель оценки уровня культуры информационной безопасности.

Для оценки уровня КИБ необходимо задать множество качественных или количественных показателей. При формировании множества показателей безопасности должны быть учтены цели безопасности G_0, факторы F_0 и задачи безопасности $Obj = \{obj_1, obj_2, ..., obj_n\}$ (модель $A_{concept}$). Для описания соответствия реального результата деятельности по ЗИ

требуемому результату $R_{required}$ введем числовую функцию соответствия на множестве результатов:

$$\rho = \rho(Y(u), R_{required}) \tag{6}$$

Фактический результат деятельность по ЗИ $Y(u) \in R_{actual}$ может быть получен как результат оценки уровня КИБ на основе множества показателей безопасности информации W. Таким образом, возникает задача разработки системы показателей безопасности, на основе которой может получен интегральный показатель $\tilde{W}$. Модель оценки уровня КИБ в общем виде представляется следующим кортежем:

$$A_{eval} = \left\langle \rho, Y(u), R_{required}, W \right\rangle \tag{7}$$

На основе полученных результатов оценки центр безопасности в случае необходимости вносит корректировки $\omega_M, \omega_G, \omega_F, \omega_a$ на необходимом уровне институционального управления деятельностью по ЗИ. Переменная ω_M отражает изменение множества мотивов центра безопасности, ω_G и ω_F отражает изменение множества целей и факторов соответственно, ω_a означает корректировки, касающиеся управляющего воздействия, направленного на агента безопасности.

Выводы

Институциональное управление деятельностью по защите информации сегодня является перспективным подходом к решению проблем управления организационными аспектами деятельности по ЗИ. Природа данного типа управления позволяет осуществлять управление как явными, так и неявными формами норм и правил безопасности, которые накладывают ограничения на деятельность сотрудников организации или мотивирует их к соблюдению политики безопасности.

Предложенная модель управления раскрывает природу поведения агента безопасности, позволяя центру безопасности вырабатывать управленческие решения с учетом гипотезы рациональности агента безопасности. Установлено, что поведение агента зависит от его типа, меры субъективного восприятия риска и правила индивидуального выбора, которым руководствуется агент безопасности.

Активными компонентами данной модели выступают центр безопасности и агент безопасности. Поведение агента безопасности, как было показано, оказывает значительное влияние на формирование КИБ. В то же время центр безопасности, как управляющий субъект, должен формировать поведение агента, учитывая его параметры в ходе осуществления необходимых управляющих воздействий.

В качестве направления дальнейших исследований представляется актуальной разработка метода численной оценки уровня КИБ, разработка множества показателей безопасности для оценки уровня КИБ и их классификация, формализация процедуры оценивания.

Список литературы: 1. *Dhillon, G. and Torkzadeh, G.* Value-focused assessment of information system security in organizations // Information Systems Journal. 2006. 16:293-314. 2. *Siponen, M.,* Oinas-Kukkonen, H.07). A review of information security issues and respective research contributions // SIGMIS Database. 2007. 38(1):60-80. 3. *Vroom, C., von Solms, R.* Towards information security behavioral compliance // Computers & Security. 2004. 23(3):191-198. 4. *Новиков. Д.А.* Институциональное управление организационными системами. – М. : ИПУ РАН, 2004. – 68 с.

Харьковский национальный университет радиоэлектроники

Поступила в редколлегию 14.09.201

УДК 681.323

А.А.ЗАМУЛА, канд. техн. наук, С.А.СИРОТА, Н.И.КОСИКОВСКАЯ

КОЛИЧЕСТВЕННАЯ ОЦЕНКА УЯЗВИМОСТЕЙ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННЫХ СИСТЕМ

В организациях часто не учитывается тот факт, что администраторы и пользователи регулярно изменяют конфигурацию информационных систем (ИС). В результате этих изменений могут появляться новые уязвимости, связанные с операционными системами (ОС) и приложениями. Кроме того, очень быстро изменяются информационные и сетевые технологиb, регулярно появляется новое ПО. Непрерывное развитие сетевых технологий при отсутствии постоянного анализа их безопасности и нехватке ресурсов для обеспечения защиты приводит к тому, что со временем защищенность ИС падает, так как появляются новые неучтенные угрозы и уязвимости системы.

В большинстве случаев для решения возникших проблем с защитой информации в организациях используются технические подходы. Администраторы безопасности имеют тенденции реагировать только на те риски информационной безопасности, которые им понятны. Фактически рисков может быть существенно больше. Только строгий постоянный контроль защищенности ИС и комплексный поход, обеспечивающий единую политику безопасности, позволяют существенно снизить риски безопасности.

Атакой на ИС является любое действие, выполняемое нарушителем для реализации угрозы путем использования уязвимостей ИС. Под уязвимостью ИС понимается любая характеристика или элемент ИС, использование которых нарушителем может привести к реализации угрозы.

Эффективное применение информационных технологий – общепризнанный стратегический фактор роста конкурентоспособности компании. Многие предприятия переходят к использованию широких возможностей Интернета и электронного бизнеса, неотъемлемым элементом которого являются электронные транзакции.

Цель статьи – оценка уязвимостей информационно-телекоммуникационных систем и классификация систем обнаружения атак на информационные системы.

Системы анализа защищенности

Для любой компании (финансовой, страховой, торговой и т.п.) существует типовая информационная система, состоящая из компонент, решающих специфичные задачи, но в общем случае ИС включает в себя четыре уровня функционирования:

1. Уровень прикладного программного обеспечения, отвечающий за взаимодействие с пользователем. Примерами элементов ИС, работающих на этом уровне, являются редактор WinWord, редактор электронных таблиц Excel, почтовая программа Outlook, системы MS Query и т.д.

2. Уровень системы управления базами данных (СУБД), отвечающий за хранение и обработку данных информационной системы. Примерами элементов ИС, работающих на этом уровне, являются СУБД Oracle, MS SQL Server, Sybase и даже MS Access.

3. Уровень операционной системы (ОС), отвечающий за обслуживание СУБД и прикладного программного обеспечения. Примерами элементов ИС, работающих на этом уровне, являются ОС MicrosoftWindowsNT, SunSolaris, NovellNetware и другие.

4. Уровень сети, отвечающий за взаимодействие узлов информационной системы. Примерами элементов ИС, работающих на этом уровне, являются стеки протоколов TCP/IP, IPS/SPX и SMB/NetBIOS.

Злоумышленники обладают широким спектром возможностей по нарушению политики безопасности. Воздействия злоумышленника могут быть осуществлены на указанных уровнях функционирования ИС.

С целью повышения уровня защищенности ИС от различных атак необходимо наряду с "традиционными" средствами защиты (например, межсетевые экраны) использовать

"адаптивные" средства: система анализа защищенности, системы обнаружения атак и другие.

Системы анализа защищенности проводят всесторонние исследования систем для обнаружения уязвимостей, которые могут привести к нарушениям политики безопасности. Результаты, полученные от средств анализа защищенности, представляют "мгновенный снимок" состояния защиты системы в данный момент времени. Несмотря на то, что эти системы не могут обнаруживать атаку в процессе ее развития, они могут определить возможность реализации атак.

При решении практических задач защиты информации первостепенное значение имеет количественная оценка ее уязвимости. Поскольку воздействие на систему различных факторов в значительной мере является случайным, то в качестве количественной меры уязвимости системы наиболее целесообразно принять вероятность нарушения защищенности, а также потенциально возможный размер ущерба, наносимого таким воздействием. При этом основными параметрами, влияющими на вероятность нарушения защищенности информации, являются: количество и типы структурных компонентов системы или объекта, количество и типы случайных угроз, количество и типы преднамеренных угроз, число и категории лиц, которые потенциально могут быть нарушителями установленных правил обработки информации, и, наконец, виды защищаемой информации.

Оценки уязвимостей ИС

Известно, что несанкционированное получение информации возможно не только путем непосредственного доступа к базам данных, но и многими другими путями, не требующими такого доступа. При этом основную опасность представляют преднамеренные действия злоумышленников. Воздействие случайных факторов само по себе не ведет к несанкционированному получению информации, оно лишь способствует появлению каналов несанкционированного получения информации (КНПИ), которыми может воспользоваться злоумышленник.

Для несанкционированного получения информации необходимо одновременное наступление следующих событий:

-- нарушитель должен получить доступ в соответствующую зону;

-- во время нахождения нарушителя в зоне в ней должен проявиться соответствующий КНПИ;

-- КНПИ должен быть доступен нарушителю соответствующей категории;

-- в момент доступа нарушителя к КНПИв данном канале должна находиться защищаемая информация.

Безусловно, важным является получение количественной оценки уязвимости информации. В качестве такой оценки может быть использована вероятность несанкционированного получения информации (P_{ijkl})нарушителем k-й категории по j-му КНПИ в l--й зоне информационных ресурсов i-го структурного компонента системы:

$$P_{ijkl} = P_{ijl}^{(d)} P_{ijl}^{(k)} P_{ijkl}^{(h)} P_{ijl}^{(u)}, \qquad (1)$$

где $P_{ikl}^{(d)}$ – вероятность доступа нарушителя k-й категории в l-ю зону информационных ресурсов i-го компонента компьютерной системы;

$P_{ijl}^{(u)} \mathbf{P_{ijl}^{(k)}}$ – вероятность наличия j-го КНПИ в l-й зоне информационных ресурсов i-го компонента компьютерной системы;

$P_{ijkl}^{(h)}$ – вероятность доступа нарушителя k-й категории к j-му КНПИ в l-й зоне информационных ресурсов i-го компонента при условии доступа нарушителя к базе данных;

$P_{ijl}^{(u)}$ – вероятность наличия защищаемой информации в j-м КНПИ в l-й зоне информационных ресурсов i-го компонента в момент доступа туда нарушителя.

Вероятность несанкционированного получения информации в одном компоненте компьютерной системы одним злоумышленником одной категории по одному КНПИ. Назовем базовым показателем уязвимости информации [2]. С учетом (1) выражение для базового показателя будет иметь вид

$$P^{б}_{ijk} = 1 - \prod_{l=1}^{5}(1 - P_{ijkl}) = 1 - \prod_{l=1}^{5}\left[1 - P^{(d)}_{ijl} P^{(k)}_{ijz} P^{(h)}_{ijkl} P^{(u)}_{ijl}\right). \quad (2)$$

Рассчитанные таким образом базовые показатели уязвимости сами по себе имеют ограниченное практическое значение. Для решения задач, связанных с разработкой и эксплуатацией систем защиты информации, необходимы значения показателей уязвимости, обобщенные по какому-либо одному индексу (i, j, k) или по их комбинации. Рассмотрим возможные подходы к определению таких частично обобщенных показателей.

Пусть {K} – интересующее нас подмножество из полного множества потенциально возможных нарушителей. Тогда вероятность нарушения защищенности информации указанным подмножеством нарушителей по *n*-му фактору в *i*-м компоненте системы ($P_{ij(K^*)}$) может быть определена следующим образом:

$$P_{ij(K^*)} = 1 - \prod_{K^*}\left[1 - P^{б}_{ijk}\right], \quad (3)$$

где Π_{K^*} означает перемножение выражений в скобках для всех нарушителей, входящих в подмножество $\{K^*\}$.

Аналогично, если $\{J^*\}$ есть подмножество представляющих интерес КНПИ, то уязвимость информации в *i-м* компоненте по данному подмножеству факторов относительно *k*-го нарушителя может быть определена следующим образом:

$$P_{i(J^*)k} = 1 - \prod_{J^*}\left[1 - P^{б}_{ijk}\right]. \quad (4)$$

Если же $\{I^*\}$ есть подмножество интересующих нас структурных компонентов системы, то уязвимость информации в них по *j*-му КНПИ относительно *k-го* нарушителя:

$$P_{(I^*)jk} = 1 - \prod_{I^*}\left|1 - P^{б}_{ijk}\right|. \quad (5)$$

Каждое из приведенных выше выражений позволяет производить обобщение по одному из множества параметров. Нетрудно получить и общее выражение, если нас интересуют подмножества $\{I\}$, $\{J\}$ и $\{K\}$ одновременно. В этом случае:

$$P_{(I^*)(J^*)(K^*)} = 1 - \prod_{I^*}\left[1 - P^{б}_{ijk}\right]\prod_{K^*}\left[1 - P^{б}_{ijk}\right]\prod_{K^*}\left[1 - P^{б}_{ijk}\right]. \quad (6)$$

Очевидно, общий показатель уязвимости P при таком подходе:

$$P = 1 - \prod_{I^*}\left[1 - P^{б}_{ijk}\right]\prod_{K^*}\left[1 - P^{б}_{ijk}\right]\prod_{K^*}\left[1 - P^{б}_{ijk}\right]. \quad (7)$$

На практике наибольший интерес представляют экстремальные показатели уязвимости, характеризующие наиболее неблагополучные условия защищенности информации: самый уязвимый структурный компонент системы (I^*), самый опасный КНПИ (J^*), самая опасная категория нарушителей (K^*).

Аналогичным образом может быть проведена оценка уязвимости информации и в других случаях, в частности в случаях нарушения целостности.

Для расчета показателей уязвимости информации с учетом интервала времени, на котором оценивается уязвимость, следует учитывать, что чем больше интервал времени, тем больше возможностей у нарушителей для злоумышленных действий и тем больше вероятность изменения состояния системы и условий обработки информации.

Можно определить такие временные интервалы (не сводимые к точке), на которых процессы, связанные с нарушением защищенности информации, являлись бы однородными. Назовем такой интервал малым. Такой малый интервал, в свою очередь, может быть разделен на очень малые интервалы, уязвимость информации на каждом из которых определяется независимо от других. При этом, в силу однородности происходящих процессов, уязвимость информации на каждом из выделенных очень малых интервалов будет определяться по одной и той же зависимости.

Тогда, если P_t^m – показатель уязвимости на фиксированном интервале, а $P^м$-показатель уязвимости на малом интервале, то:

$$P^м = 1 - \prod_{t=1}^{n} [1 - P_t^m], \tag{8}$$

где t – интервал деления выбранного объекта СУБД интервалов наотдельные сегменты; n – общее число очень малых интервалов.

Рассмотренный подход можно распространить и на другие интервалы, а именно: большой интервал представить некоторой последовательностью малых, очень большой – последовательностью больших, бесконечно большой – последовательностью очень больших интервалов [2].

Системы обнаружение атак

Системы обнаружения атак реализуются посредством анализа журналов регистрации операционной системы и прикладного ПО, сетевого трафика в реальном времени. Компоненты обнаружения атак, размещенные на узлах или сегментах сети, оценивают различные действия, в том числе, использующие известные уязвимости [3].

Рассмотрим этапы осуществления атаки. Первый, подготовительный этап, заключается в поиске злоумышленником предпосылок для осуществления той или иной атаки. На этом этапе злоумышленник ищет уязвимости в системе.

На втором, основном этапе реализации атаки, – осуществляется использование найденных уязвимостей. На третьем, заключительном этапе, злоумышленник завершает атаку и старается скрыть следы вторжения [1].

Существующие механизмы защиты, реализованные в межсетевых экранах (firewall), серверах аутентификации, системах разграничения доступа и т.д., работают только на втором этапе. Т.е. по существу они являются средствами блокирующими, а не упреждающими атаки. Комплексная система обеспечения информационной безопасности должна работать на всех трех этапах осуществления атаки. И обеспечение адекватной защиты на третьем, завершающем, этапе не менее важно, чем на первых двух. Ведь только в этом случае можно реально оценить ущерб от "успешной" атаки, а также разработать меры по устранению дальнейших попыток реализовать аналогичную атаку.

Обнаруживать, блокировать и предотвращать атаки можно несколькими путями. Первый, самый распространенный, способ – обнаружение уже реализуемых атак. Этот способ применяется в "классических" системах обнаружения атак (например, RealSecure компании Internet Security Systems), межсетевых экранах и т.п. Однако "недостаток" средств данного класса в том, что атаки могут быть реализованы повторно. Они также повторно обнаруживаются и блокируются. Второй путь – предотвратить атаки еще до их реализации. Осуществляется это поиском уязвимостей, которые могут быть использованы для реализации атаки. И, наконец, третий путь – обнаружение уже совершенных атак и предотвращение их повтор-

юго осуществления. Таким образом, системы обнаружения атак могут быть классифицированы по этапам осуществления атаки (рис. 1):

- системы, функционирующие на первом этапе атак и позволяющие обнаружить уязвимости информационной системы, используемые нарушителем для реализации атаки. Иначе средства этой категории называются системами анализа защищенности (security assessment systems) или сканерами безопасности (security scanners);

- системы, функционирующие на втором этапе атаки и позволяющие обнаружить атаки в процессе их реализации, т.е. в режиме реального (или близкого к реальному) времени. Именно эти средства и принято считать системами обнаружения атак в классическом понимании. Помимо этого в последнее время выделяется новый класс средств обнаружения атак – обманные системы;

- системы, функционирующие на третьем этапе атаки и позволяющие обнаружить уже совершенные атаки. Эти системы делятся на два класса – системы контроля целостности, обнаруживающие изменения контролируемых ресурсов, и системы анализа журналов регистрации.

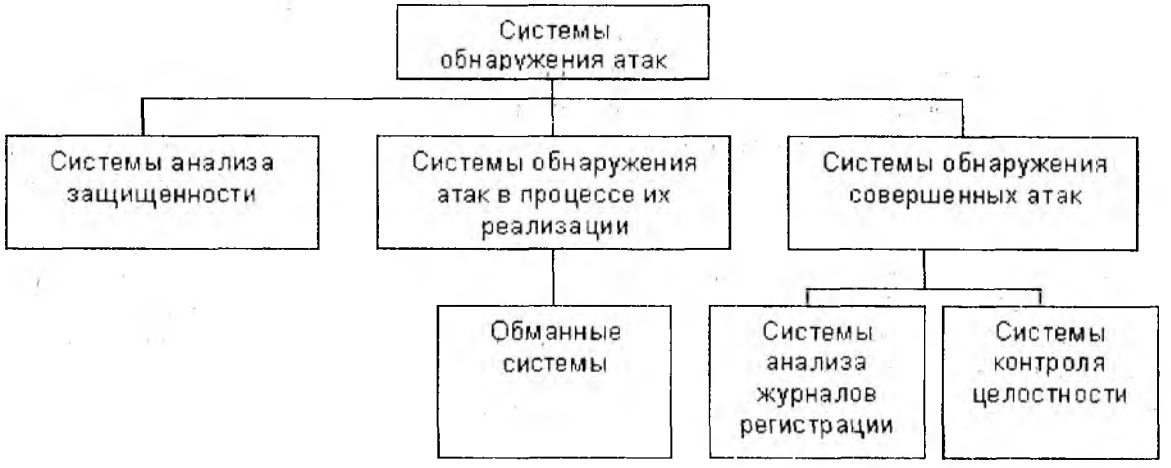


Рис.1

Помимо этого, существует еще одна распространенная классификация систем обнаружения нарушения политики безопасности – по принципу реализации: host-based, т.е. обнаруживающие атаки, направленные на конкретный узел сети, и network-based, направленные на всю сеть или сегмент сети. Обычно на этом дальнейшая классификация останавливается. Однако системы класса host-based можно разделить еще на три подуровня:

- Application IDS, обнаруживающие атаки на конкретные приложения;
- OS IDS, обнаруживающие атаки на операционные системы;
- DBMS IDS, обнаруживающие атаки на системы управления базами данных.

Выделение обнаружения атак на системы управления базами данных (СУБД) в отдельную категорию связано с тем, что современные СУБД уже вышли из разряда обычных приложений и по многим своим характеристикам, в т.ч. и по сложности, приближаются к операционным системам. Таким образом, классификация систем обнаружения атак по уровню реализации выглядит следующим образом (рис. 2). Можно заметить, что это деление соответствует уровням информационной системы предприятия .

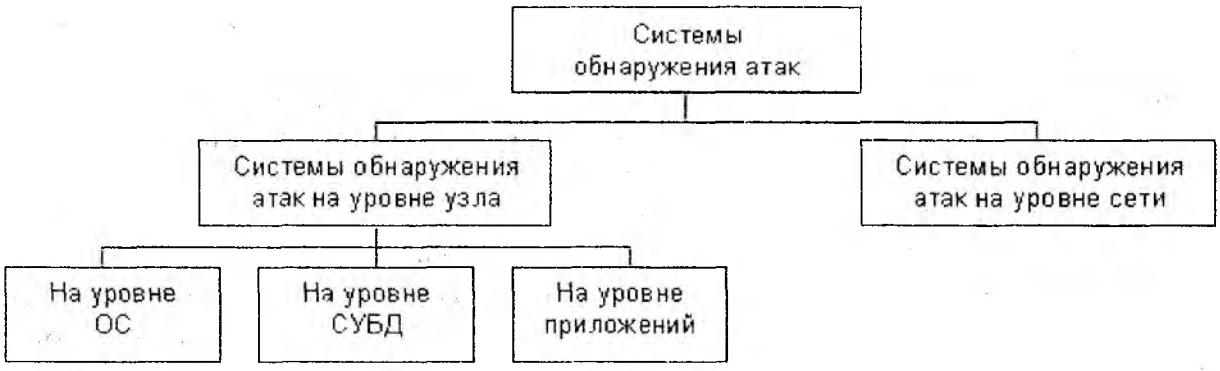


Рис. 2

Выводы

1. Таким образом, применение "традиционных" средств безопасности не позволяе реализовать эффективную защиту от способов воздействия на информационные ресурсы компании. Только комплексное применение "традиционных" и "адаптивных" механизмо позволит гарантировать высокий уровень защищенности информационной системы предприятия от внешних и внутренних злоумышленников.

2. Для того чтобы привести систему обеспечения информационной безопасности организации в соответствие современным требованиям, необходимо дополнить имеющееся решения компонентами, реализующими анализ защищенности ИС и обнаружение атак на ИС.

Список литературы: *1. Малюк, А.А.* Информационная безопасность. – М. : Горячая линия Телеком 2004. – 280с. 2. *Шаньгин, В.Ф.* Информационная безопасность компьютерных систем и сетей. – М. ИД «ФОРУМ», 2008. – 415с. 3. *Лукаций, А.В.*Обнаружение атак. – СПб. : БХВ-Петербург, 2004. - 624с. 4. *Лукаций, А.В.* Безопасность беспроводных сетей // Технологии и средства связи. – 2005. – № 5. *Зима, В.М.* Компьютерные сети и защита передаваемой информации. – СПб. : Изд. СПбГУ, 2003. - 569с. 6. *Костров, Д.* Системы обнаружения атак // BYTE Россия. – 2002. – №8

Харьковский национальный университет радиоэлектроники

Поступила в редколлегию 13.09.201

УДК 621.391

А. А. ЗАМУЛА, канд. техн. наук

ПРЕДЛОЖЕНИЯ ПО ПОСТРОЕНИЮ ШИРОКОПОЛОСНЫХ СИСТЕМ ПЕРЕДАЧИ СО СЛОЖНЫМИ СИГНАЛАМИ

Введение

Одним из методов обмена данными в системах передачи информации (СПД) является метод, основанный на использовании динамической смены соответствия: *m* бит сообщения – 2^m сложных сигналов. Для реализации метода должны быть определены требования к источнику сложных сигналов с точки зрения решения классических задач теории оптимального приема (обнаружение, различение, оценка параметров и др.), а также сформулированы условия, обеспечивающие установление такого правила смены соответствия *m* бит сообщения – 2^m сложных сигналов, предсказание которого нарушителем возможно с вероятностью, не превышающей допустимую.

Цель статьи – разработка предложений по построению скрытной системы передачи информации на основе использования сигналов с расширенным спектром.

Постановка задачи выбора сложных сигналов для широкополосных систем передачи

Известно, что любой сигнал $S(t)$ конечной энергии может быть представлен как сумма несчётного числа гармонических колебаний, амплитуды и фазы которых в пределах бесконечно малого диапазона частот $[f, f+df]$ определяются спектральной плотностью или спектром $\overline{S}(f)$. Математическим отображением этого факта служит пара обратного и прямого преобразования Фурье

$$S(t)=\int_{-\infty}^{\infty}\overline{S}(f)\exp(j2\pi ft)df,\ \widetilde{S}(f)=\int_{-\infty}^{\infty}S(t)\exp(-2\pi ft)dt\,. \qquad (1)$$

Для характеристики размера зоны, в которой сосредоточена энергия сигнала во временной или частотной области, используются обозначения длительности сигнала T или полосы ΔF соответственно. Детерминированный сигнал, для которого $T\Delta F >> 1$, и полоса которого может изменяться независимо от длительности, называется сигналом с расширенным спектром или широкополосным.

Типичным для теории связи является подход, заключающийся в разработке оптимального приемного устройства, которое с наилучшим качеством восстановит информацию, содержащуюся в наблюдаемом колебании. Определение оптимального алгоритма обработки, базирующегося на учёте специфических свойств переданного сигнала, позволяет синтезировать оптимальным образом и сам сигнал, т.е. выбрать наилучшим образом метод его кодирования и модуляции.

В теории связи наиболее распространённой моделью служит канал с аддитивным белым гауссовским шумом, в котором вероятность трансформации каналом заданного входного сигнала в то или иное выходное наблюдение $y(t)$ (переходная вероятность – $P[y(t)\,|\,S(t)]$) экспоненциально уменьшается с ростом квадрата Евклидова расстояния между переданным сигналом и выходным наблюдением:

$$P\left[y(t)\middle|S(t)\right]=\kappa\exp\left(-\frac{1}{N_0}\,d(s,y)\right), \qquad (2)$$

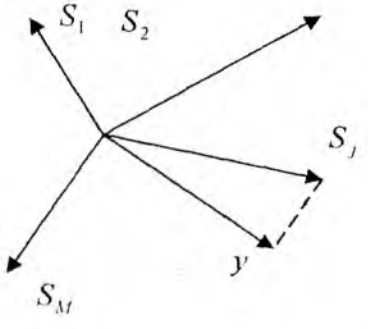
где κ – константа, не зависящая от $S(t)$ и $y(t)$, N_0 – спектральная плотность мощности одностороннего белого шума; а Евклидово расстояние между $S(t)$ и $y(t)$ определяется как

$$d(S,y)=\sqrt{\int_0^T [y(t)-S(t)]^2 dt}. \qquad (3)$$

Согласно соотношениям (2) и (3) похожесть сигнала (вероятность того, что он преобразован каналом в наблюдение $y(t)$ уменьшается с увеличением Евклидова расстояния между $S(t)$ и $y(t)$. В случае равной вероятности всех сообщений источника (что достигается при правильном проектировании системы) оптимальной стратегией наблюдателя, обеспечивающей минимальную ошибку перепутывания действительно переданного с некоторым другим сигналом, является правило (критерий) максимального правдоподобия (МП). Согласно данному алгоритму, после того, как колебание $y(t)$ стало достижимым (принято), решение принимается в пользу того сигнала, для которого вероятность трансформации его каналом в принятое наблюдение $y(t)$ является наибольшим (по сравнению с вероятностями для других сигналов). С учётом изложенного, МП решение для гауссова канала может быть преобразовано в правило минимума расстояния

$$d(S_j,y)=\min d(S_j,y) \Rightarrow H_j, \qquad (4)$$

т.е. решение принимается в пользу сигнала $S_j(t)$, поскольку он наиболее близок (в смысле Евклидова расстояния) к наблюдению $y(t)$ среди всех конкурирующих сигналов (рис. 1).



$$d(S_j,y)=\min d(S_i,y)$$

Рис. 1

С учётом используемой геометрической интерпретации сигналов можно ввести длину сигнала $\|S\|$ как его расстояние относительно начала координат. Тогда из (3) следует, что $\|S\|=d(S,0)=\sqrt{E}$, где

$$E=\int_0^T S^2(t)dt, \qquad (5$$

– энергия сигнала. Другой важной геометрической характеристикой является скалярное произведение (U,V) двух сигналов: $U(t),V(t)$

$$(U,V)=\int_0^T U(t)\cdot V(t)dt, \qquad (6$$

которое может трактоваться как предельная форма скалярного произведения двух n-мерных векторов. Эта же характеристика может быть вычислена с помощью длины векторов и косинуса угла α между ними: $(U,V)=\|U\|\|V\|\cos\alpha$, и, таким образом, скалярное произведение свидетельствует о близости или похожести сигналов, поскольку, чем ближе сигналы одинаковой длины (энергии) друг к другу, тем меньше $\cos\alpha$ отличается от единицы, и тем больше скалярное произведение. На основании этого скалярное произведение (6) называют также корреляцией сигналов.

Раскрыв скобки в (3), приходим к соотношению

$$d^2(S_i,y)=\int_0^T y^2(t)dt-2\int_0^T y(t)\cdot S(t)dt+\int_0^T S^2(t)dt=\|y\|^2-2Z_i+\|S_i\|^2, \qquad (7)$$

где Z_i – соответствует корреляции между наблюдением $y(t)$ и i-м сигналом $S_i(t)$:

$$Z_i(y,S_i)=\int_0^T y(t)\cdot S(t)dt. \qquad (8)$$

Первое слагаемое в правой части соотношения (7) фиксировано для данного наблюдения и не влияет на анализируемые расстояния и решение, какой из сигналов был принят. Последний член суммы есть ни что иное, как энергия i-го сигнала E_i. Учитывая это, правило минимума расстояния (4) может быть сформулировано как правило максимума корреляции:

$$Z_j-\frac{E_j}{2}=\max(Z_i-\frac{E_j}{2})\Rightarrow H_j, \qquad (9)$$

означающее, в частности, что из M возможных сигналов с одинаковой энергией фактически принятым считается тот, который имеет максимум корреляции с наблюдением $y(t)$.

Приведенные рассуждения указывают на способ конструирования множества сигналов. На рис. 2 изображены сигнальные векторы.

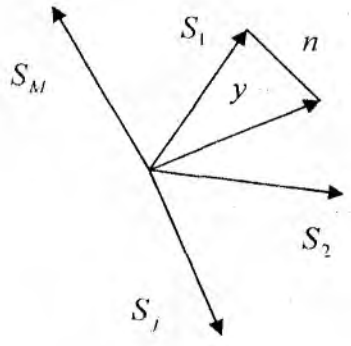


Рис. 2

Предположим, что передавался сигнал S_1 и что он подвергается искажению в канале с аддитивным белым гауссовым шумом, следствием чего служит добавление к S_1 вектора шума n. Гауссовский вектор n характеризуется симметричным (сферическим) вероятностным распределением, экспоненциально спадающим с увеличением длины вектора n, что очевидно следует из (2) после удаления из него сигнала (т.е. при подстановки $S(t)=0$). Следователь-

но, вектор наблюдения $y = s_1 + n$ будет случайным образом перемещаться вокруг S_1, как это показано на рисунке, и тогда, согласно правилу минимума расстояния (4), как только y окажется ближе к некоторому другому, чем S_1 сигналу, то будет принято ошибочное решение. Для минимизации вероятности возникновения такого рода ошибки следует располагать другие сигналы на максимально большом расстоянии от S_1. Поскольку любой из M сигналов может передаваться равновероятно, т.е. занимать место S_1, то, очевидно, что все расстояния следует делать максимально большими.

Задача построения множества максимально удалённых друг от друга сигналов (входящая в класс так называемых задач упаковки) оказывается достаточно сложной и пока что не имеет общего решения.

Различают две классические версии широкополосной модуляции: прямой последовательностью и $d(S_i, S_j), 1 \le i \le j \le M$ прыгающей частотой [1]. Идея прямого расширения спектра состоит в амплитудно-фазовой модуляции дискретных последовательностей (сигнатур) пользователя потоком данных. Например, выполняется умножение битового информационного потока $B_k(t)$-го абонента на специфическую для каждого пользователя (в многопользовательских системах, например в CDMA приложениях) на сигнатуру $S_k(t)$, а результат произведения $S_k(t) \cdot B_k(t)$ моделирует непрерывную несущую, т.е.

$$S_k(t, b_k) = S_k(t) \cdot B_k(t) \cdot \cos(2\pi f_0 t)$$

где $b_k = (..., b_{k,-1}, b_{k,0}, b_{k,1}, ...)$ – битовый поток k-го пользователя, а $B_k(t) = b_{k,i} = \pm 1, (i-1)T_b < t < iT_b$ (T_b – длительность импульсов положительной и отрицательной полярности информационного сигнала k-го пользователя).

В описанном примере битовый поток первоначально модулирует бинарную сигнатуру, а результат используется для бинарной фазовой манипуляции несущей.

В [1] показано, что в качестве сигнатур в системе с прямым расширением должен использоваться такой ансамбль детерминированных последовательностей, все представители которого в идеале обладают нулевой постоянной составляющей, идеальной периодической автокорреляционной функцией (АКФ) и нулевой периодической взаимокорреляционной функцией (ВКФ).

Широкое практическое использование имеют минимаксные ансамбли: m-последовательности, множества Голда, множества Кассами и их расширения, ансамбли Камалетдинова, характеристические коды и др.

Условия построения абсолютно стойкой системы передачи на уровне источника сигналов

Обоснование выбора того или иного класса минимаксных ансамблей для рассматриваемого метода передачи данных выходит за рамки данной статьи. Основное внимание будет сосредоточено на условиях, обеспечивающих установление правила смены соответствия m бит сообщения – $2m$ сложных сигналов, при выполнении которых предсказание нарушителем правила смены соответствия возможно с вероятностью, не превышающей допустимую. И в этом смысле можно говорить о некотором значении криптографической стойкости СПИ на уровне источника сложных сигналов.

На рис. 3 представлена структурная схема формирования сигнала в динамическом радиоканале, состоящая из ИИ – источника информации, ДМ – динамического модулятора, УФУП – устройства формирования управляющей последовательности.

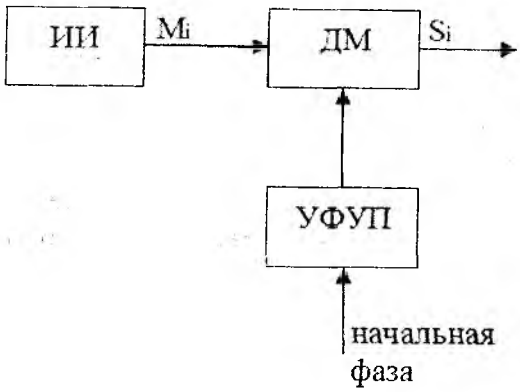


Рис. 3

Пусть имеется некоторый источник информации, создающий в фиксированный момент времени одно из M возможных сообщений. Каждое из M конкурирующих сообщений передается посредством сигнала: $S = \{S_i(t): i = 1,2, \ldots, M\}$ вида (1). На число сигналов M (или мощность множества S) не накладываются никакие ограничения и, если это необходимо, множество S может быть бесконечным.

Радиоканалы, в которых выполняются эти условия, называют динамическими.

Модулятор обеспечивает формирование сложных сигналов, а демодулятор (устройство обработки сложных сигналов) – их поиск, обнаружение и различение. Символы сообщения от источника информации, представленные в виде m бит, поступают в динамический модулятор, в котором в соответствии с символами управляющей последовательности УФУП, осуществляется выбор 2^m из M сложных сигналов и таким образом устанавливается соответствие: m бит – 2^m сложных сигналов. При появлении на входе динамического модулятора m бит сообщения в канал связи излучается сложный сигнал S_i, выбранный в зависимости от значения управляющей последовательности.

По истечении времени T соответствие m бит – 2^m сложных сигналов изменяется по определенному закону (правилу).

Нарушитель, осуществляющий наблюдение за каналом связи, может реализовывать различные стратегии воздействия на СПД: перехват излученных сигналов, их анализ, попытки распознавания сигналов и определения закона их излучения, формирование и постановка помех с целью навязывания ложных сообщений и др.

В демодуляторе на станции приема производится различение одного из 2^m разрешенных информационных сигналов. После демодуляции в соответствии решением, принятым в соответствии с (10) на выходе демодулятора формируются m бит сообщения, которые поступают получателю сообщений.

Естественным представляется постановка ряда вопросов, на которые необходимо дать ответы, если мы хотим построить СПИ, реализующую обсуждаемый метод передачи данных. Насколько устойчива система против раскрытия закона установления соответствия m бит сообщения – 2^m сложных сигналов, если нарушитель не ограничен временем и обладает всеми необходимыми средствами для анализа перехваченных сигналов. Имеет ли правило соответствия, которое определяет нарушитель, единственное решение (j – вариант соответствия: m бит – сложный сигнал), и если нет, то сколько приемлемых решений возможно. Какой объем данных (число элементов физического носителя информации, – сигнала) необходимо перехватить нарушителю, для того, чтобы решение стало единственно верным. Существуют ли системы, в которых вообще нельзя принять единственное правильное решение независимо от того, каков объем перехваченного в канале наблюдения.

Предположим, что имеется конечное число возможных дискретных сообщений $M_1, M_2, \ldots, M_N$ с априорными вероятностями $P(M_1), P(M_2), \ldots, P(M_N)$ и что эти сообщения

преобразуются в возможные сложные сигналы $S_1, S_2, \ldots S_N$. После того как нарушитель перехватил некоторый сигнал, он имеет возможность вычислить апостериорные вероятности сообщений, содержащихся в принятом наблюдении.

Сформулируем необходимые и достаточные условия построения абсолютно стойкой системы на уровне источника сигналов.

Пусть каждые m бит источника сообщений в интервале T ставятся в соответствие 2^m сложных сигналов S, выбранных из пространства {S} размерности $N \geq 2$, тогда необходимыми и достаточными условиями абсолютной стойкости на уровне источника сигналов являются условия [1]

$$P(M_i / S_i) = (M_i); \tag{10}$$

$$P(S_i / S_{i-1}, S_{i-2}, S_v, \ldots, S_1) = P(S_i). \tag{11}$$

Другими словами: вероятность появления S_i сигнала в канале не должна зависеть от того, какое сообщение появилось на выходе источника; вероятность появления S_i сигнала в канале не должна зависеть от того, какие сигналы до этого излучались. В этом случае перехват нарушителем сигнала не дает ему никакой информации, необходимой для определения содержания сообщения, содержащегося в полученных сигналах. С другой стороны, если эти условия равенства вероятностей не выполнены, то имеют место случаи, когда для определенного варианта соответствия m бит сообщения – 2^m сложных сигналов апостериорные вероятности, вычисленные нарушителем, отличаются от априорных. А это, в свою очередь, может повлиять на выбор нарушителем своих действий и, таким образом, не обеспечит абсолютной стойкости системы.

Необходимое условие следует из теоремы Байеса, в соответствии с которой [1]

$$P(M_i / S_i) = \frac{P(M_i) P(S_i / M_i)}{P(S_i)}, \tag{12}$$

где $P(M_i)$ – априорная вероятность (передачи сообщения M_i); $P(M_i / S_i)$ – апостериорная вероятность сообщения M_i при условии, что перехвачен сигнал S_i; $P(S_i / M_i)$ – условная вероятность сигнала S_i при условии, что выбрано сообщение M_i; $P(S_i)$ – вероятность получения сигнала S_i.

Для обеспечения абсолютной стойкости величины должно быть выполнено одно из равенств: или $P(M_i) = 0$ (это решение должно быть отброшено, так как требуется, чтобы равенство осуществлялось при значениях $P(M)$ или же $P(E_i / M_i) = P(E_i)$ для любых M и E, или $P(S_i / M_i) = P(S_i)$, тогда $P(M_i / S_i) = P(M_i)$ и система обладает абсолютной стойкостью. Тогда количество информации, содержащейся у нарушителя после перехвата сигнала I(E/M),

$$I(S, M) = H(M) - H(M / S) = H(M) - H(M) = 0, \tag{13}$$

где $H(M / S)$ – энтропия источника сообщения, при условиях, что перехвачен сигнал S; $H(M)$ – энтропия источника открытого сообщения.

Условие (10) является достаточным условием абсолютной стойкости. В этом случае определение j-варианта соответствия может быть выполнено только методом статистического опробования всевозможных вариантов, т. е. методом перебора.

Как следует из (11), вероятность появления S_i сигнала не зависит от вероятности появления всех i-1 сигналов. В этом случае количество информации в сигнале S_i после перехвата всех i-1 сигналов

$$I(S_i, S_v) = H(S_i) - H(S_i / S_v) = H(S_i) - H(S_v), \quad (14)$$

где $v = 1, i-1$.

Из выражения (11) также следует равновероятность появление сигналов, (т.е. равновероятность отображения m – бит – S_i сигнал), поэтому

$$H(S_i) = H(S_v). \quad (15)$$

Подставляя выражение (15) в выражение (14), получаем $I(S_i, S_v) = 0$.

Условие (11) является также и достаточным, так как независимость и равновероятность появления сигналов означает и равновероятность появления управляющей гаммы, символы которой статистически независимы и асимптотически равновероятны.

Сформулируем необходимое и достаточное условия абсолютной стойкости сигналов источника сложных ФМ ШПС сигналов {S}. Под абсолютной стойкостью будем понимать их идеальную структурную скрытность.

Пусть {S} есть ансамбль ФМ ШПС сигналов объема N с числом разрядов L в каждом из них, тогда для обеспечения абсолютной стойкости каждого из $S_i \in \{S\}$ сложных ФМ ШПС необходимо и достаточно, чтобы

$$P(S_{j,i} / S_{v,R}) = P(S_{j,i}), v = \overline{1, L}, R = \overline{1, N}, \quad (16)$$

т. е. чтобы вероятность появления элемента $S_{j,i}$ сложного ФМ ШПС сигнала не зависела ни от элементов ранее переданных сигналов, ни от элементов $S_{j,i-1}, S_{j,i-2}, S_{j,v}, \ldots, S_{j,2}, S_{j,1}$ сигнала S_j.

Необходимость условия (16) следует непосредственно из критерия абсолютной структурной скрытности сигнала [3]

$$S_c = \frac{l}{L}, \quad (17)$$

где l – число символов сложного сигнала, которые необходимо знать для определения закона формирования $L - l$ оставшихся.

Для случая,когда $l = L$

$$S_c = \frac{l}{L} = \frac{L}{L} = 1.$$

Поэтому по любому числу перехваченных символов $l < L$ S_j сигнала нельзя предсказать последующие $L - l$ символов как S_j сигнала, так и всех $S_v, v = \overline{1, i-1}, v \neq j$. Условие (16) является и достаточным.

Действительно, условная энтропия относительно закона формирования S_j сигнала после перехвата не менее R символов в v сигналах

$$H(S_{j,l} / S_{v,k}) = -\sum_{i=1}^{N} P(S_{j,i} / S_{v,k}) \log P(S_{j,i} / S_{v,k}), \quad (18)$$

и среднее значение условной энтропии об источнике (законе формирования) сигналов [2]

$$H\{S\}/\{S_v\} = -\sum_{j=1}^{v+1} \sum_{i=1}^{k} P(S_{j,i} / S_{v,k}) \log P(S_{j,i} / S_{v,k}), \quad (19)$$

вследствие справедливости (16), совпадает с априорной неопределенностью H({S}) источника сигналов. Поэтому количество информации, получаемое нарушителем при анализе (раскрытии закона формирования сигналов)

$$I(\{S\}/\{S_v\}) = H(\{S\}) - S(\{S\}/\{S_v\}) = H(\{S\}) - H(\{S\}) = 0 .$$

Выводы

Анализ выражений (16) – (19) дает возможность сформулировать предложения по построению скрытной СПИ на основе использования сигналов с расширенным спектром:

1. Закон формирования каждого из сигналов должен быть случайным, причем даже при перехвате L-1 из L символов сигнала не должно существовать единственного решения относительно закона его формирования.

2. Абсолютно стойким, с точки зрения закона формирования, является источник сигналов со случайным формированием всех сигналов, так как только в этом случае у нарушителя отсутствует возможность принятия правильного решения относительно используемого в системе сигнала.

Список литературы: 1. *Ipatov, Valery P.* Spread Spectrum and CDMA. Principles and Applications / Valery P. Ipatov. University of Turku, Finland and St. Petersburg Electrotechnical University 'LETI', Russia. John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, 2005. – 385 p. 2. Shannon C, Communication theory of secrecy system, Bell System Techn.J., 28, №4 (1949). 3. *Горбенко, І.Д., Горбенко, Ю.І.* Прикладна криптологія: Теорія. Практика. Застосування : монографія. Вид. 2-ге, перероб. і доп. – Харків : Форт, 2012. – 880 с. 4. *Помехозащищенность* радиосистем со сложными сигналами / Г. И. Тузова, В. А. Сивов и др ; под ред. Г. И. Тузова. – М. : Радиосвязь, 1985. – 264 с.

Харьковский национальный университет радиоэлектроники

Поступила в редколлегию 10.09.2012

УДК 004.056.5: 004.775

Б.А. БАБИЧ, А.В.САГУН. канд. техн. наук, О.А. КОЖУХОВСКАЯ. канд. техн. наук, А.Д. КОЖУХОВСКИЙ, д-р техн. наук

СИНТЕЗ СИСТЕМЫ МНОГОУРОВНЕВОЙ ЗАЩИТЫ КОРПОРАТИВНЫХ ПОРТАЛОВ НА ПЛАТФОРМЕ MS SHAREPOINT

Введение

Для IT, развивающихся в режиме нон-стоп, очень характерна проблема определения лидирующей платформы корпоративного портала, когда маркетинговая информация о продукте или технологии преобладает над какой-либо другой. Эта проблема особенно остро ощущается в сегменте программного обеспечения (ПО) для бизнеса. Яркий пример – *корпоративный портал* [1].

Под корпоративным порталом будем понимать видимую для пользователей внутреннюю часть сети организации. Корпоративный портал может включать в свой состав корпоративный сайт либо персональные сайты для сотрудников компании, может быть частью корпоративной информационной системы (КИС) и, собственно, КИС, обеспечивая одновременно единый доступ к информации и приложениям компании, совместную работу и управление знаниями, защиту и конфиденциальность коммерческой информации и коммерческой тайны [2].

Платформа MS Share Point представляет собой лидирующий на рынке продукт, который обеспечивает всё необходимое для развитого корпоративного портала – набор технологий, основанных на серверной операционной системе MS Windows Server, Internet Information Server, Windows Share Point Services. В качестве хранилища данных, предоставляющего возможность масштабируемости и надежности, используется MS SQL Server.

Цель исследования – изучение вопроса построения многоуровневой системы защиты для корпоративных порталов на платформе MS Share Point при помощи интегрируемого программного обеспечения Microsoft; определение алгоритма построения защищенной облачной системы универсального типа.

Моделирование защищенного корпоративного портала

На базе универсальной топологии Share Point, которая сочетает три различные архитектуры (IIS, .NET и SQL Server) [2] разработана оптимальная топология корпоративного портала с наиболее полным функциональным набором компонент [3] (рис. 1), которая может обеспечить выполнение задач сотрудников компании среднего бизнеса с учётом требований защиты информации.

Режим изоляции рабочих процессов используется в IIS по умолчанию. Режим изоляции рабочих процессов позволяет использовать преимущества архитектуры IIS 6.0 при определении групповых политик безопасности: устойчивая работа приложений в группах; учитывать автоматические перезапуски, масштабируемость, отладку и точную настройку производительности системы. Веб-приложения выполняются с удостоверением сетевой службы, которое обеспечивает повышенную безопасность в рамках субъект-субъективной модели доступа: учетная запись сетевой службы имеет меньшие права доступа, чем учетная запись локального компьютера. Актуальность включения данного режима особенно актуальна при синтезе систем многоуровневой защиты корпоративных порталов на базе продуктов Microsoft. Такой режим нужно использовать при изоляции рабочих процессов, если не требуется запуск приложений, которые могут конфликтовать с ним, что характерно для защиты корпоративных порталов.

Таким образом, определены основные узлы системы:

- Front-end Web Server (Share Point server 2010; обработка запросов пользователей);
- Application Server (Share Point server 2010, средства аналитики; обработка запросов пользователей);
- SQL Server (Share Point Data Bases, Data Warehouses, средства аналитики; хранение информации, статистическая обработка);
- *Domain Controller Server (Active Directory; учет пользователей Share Point фермы).*

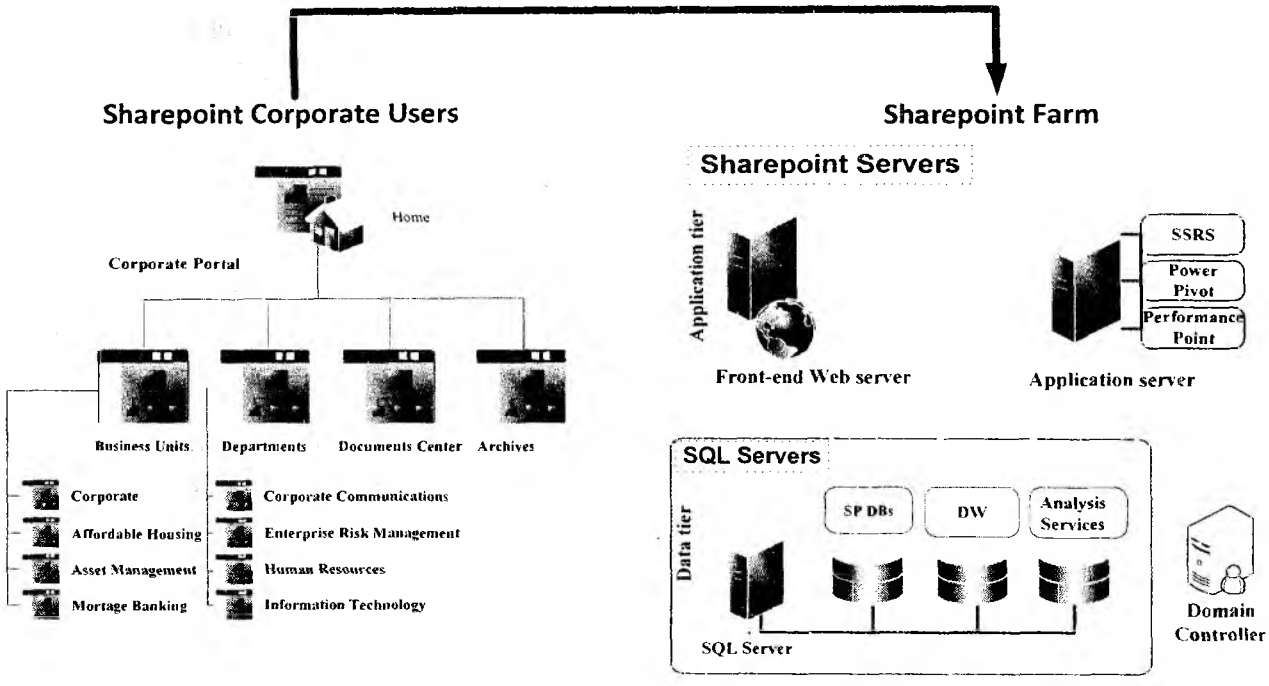


Рис. 1. Функциональная схема взаимодействия и обработки задач клиентов предложенной конфигурации на базе универсальной топологии Share Point Farm

В результате практической реализации, предложенной на рис. 1, конфигурации развернут тестовый портал, который может успешно масштабироваться и обрабатывать задачи пользователей корпораций среднего и большого бизнеса. Позволяет гибкую настройку рабочих областей для каждой рабочей группы и разделяемых ресурсов.

Анализ многоуровневой защиты предложенной системы

Рассматривая вопросы безопасности облачных систем, к которым относится данная разработка, необходимо проанализировать методы и принципы, которые предоставят пользователям системы максимальный уровень защиты для их персональных данных. Представленная система корпоративного облачного портала, согласно НД ТЗІ 2.5-005 -99, относится к А класса 3 с повышенными требованиями к обеспечению конфиденциальности, целостности доступности обрабатываемой информации [4]. Принципы защиты подобных систем определяются мощностью средств контроля и зависят от конфиденциальности ресурса. Одной из главных особенностей систем данного вида является то, что большинство ресурсов могут быть виртуализированы.

Таким образом, клиенты, пользующиеся службами, размещенными в облаке, могут иметь ресурсы, которые невозможно связать с физическим объектом – данные могут храниться виртуально и распределяться по нескольким местоположениям.

При внедрении системы было выяснено, что данный факт приводит к изменениям анализа рисков и применения элементов контроля безопасности к традиционным уровням многоуровневой защиты: оборудование, сеть, доступ к удостоверениям, авторизация доступа и аутентификация, хостинг.

Принцип многоуровневой защиты является базовым элементом в предоставлении защищенной облачной инфраструктуры. Применение средств контроля на нескольких уровнях подразумевает задействование механизмов защиты, разработку стратегий по снижению риска и способность реагировать на атаки в случае их возникновения. Использование комплекса мер обеспечения безопасности различной силы (в зависимости от конфиденциальности защищаемого ресурса) приводит к повышению эффективности предотвращения проникновения в систему и снижению негативного влияния инцидента безопасности.

Разумеется, при этом по-прежнему необходимо предпринимать меры обеспечения физической и сетевой безопасности. Однако важнейший момент управления рисками смещается в сторону уровня объекта и элементов, используемых в облачной среде: контейнеров хранилищ статических или динамических данных, объектов виртуальных машин, сред выполнения, в которых производятся вычисления.

Корпорация Microsoft обеспечивает определение внешнего и внутреннего периметров и усиление средств защиты на каждом слое периметра [5, 6]. На рис. 2 отображены продукты Microsoft, которые способны предоставить безопасность для центров обработки данных, сетевого оборудования и средств связи представленной на рис. 2 системы в следующем:

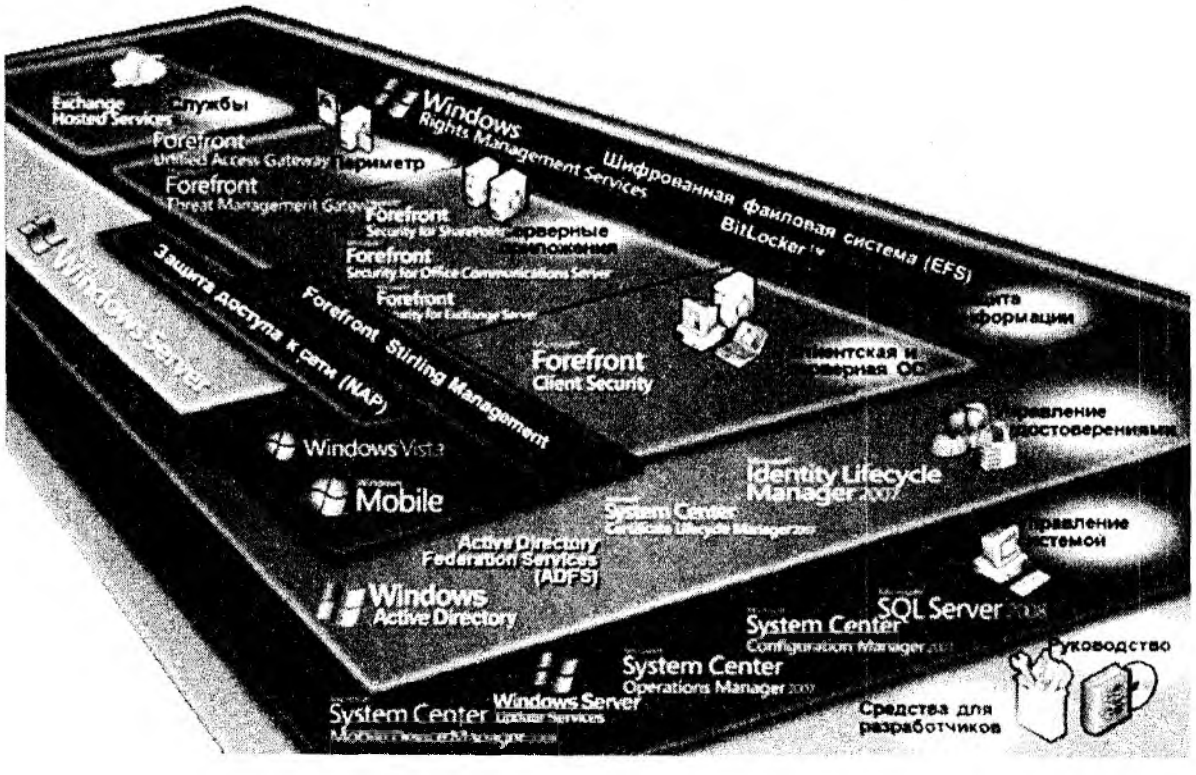


Рис. 2. Программные продукты для обеспечения многоуровневой защиты

защита клиентской и серверной операционных систем;
управление удостоверениями;
управление системой;
защита доступа к сети передачи данных;
контроль работы служб и серверных приложений.

В представленной разработке было применено несколько уровней безопасности для устройств центров обработки данных и сетевых подключений. Например, элементы контроля безопасности используются как для контроля, так и для управления. Имеется специализированное оборудование (например, балансировку нагрузки выполняет, параллельно своим основным задачам, Web-Frontend) для управления потоками данных.

Заключение

Упомянутые принципы обеспечения безопасности для сред облачных вычислений, а именно координированное и стратегическое их использование для процессов и технологий, позволяет облачной инфраструктуре адаптироваться к стремительным изменениям в сфере IT-технологий в рамках построения защищенных информационных систем.

Посредством этих программных продуктов возможно создать комплексную инфраструктуру обеспечения и контроля безопасности, обеспечивающих необходимый уровень надежности, ожидаемый заказчиками, и соответствие действующему международному стандарту ISO/IEC 27001:2005 [7] с повышенными требованиями к обеспечению конфиденциальности, целостности и доступности обрабатываемой информации.

Список литературы: 1. *Стратегия* построения защищенных информационных систем (Trustworthy Computing) корпорации Microsoft: http://www.microsoft.com/twc. 2. *Уведомление* корпорации Microsoft о конфиденциальности в сети: http://www.microsoft.com/privacy. 3. *Сертификат* ISO 27001: 2005 группы Microsoft Global Foundation Services: http://www.bsiglobal. com/en/Assessment-and-certification-cervices/Client-directory/CertificateClient-Directory-Search-Results/?pg=1&licencennumber= IS+533913& searchkey=companyXegXmicrosoft. 4. *Microsoft* Clobal Foundation Services:http://www.globalfoundation services.com. 5. *Процесс* Microsoft Security Development Lifecycle (SDL): http://www.microsoft.com/en us/library/cc307748.aspx. 6. *Центр* Microsoft Security Response Center: http://www.microsoft. com/security/msrc. 7. *Службы* Microsoft Online: http://www.microsoft.com/online

Черкасский государственный технологический университет

Поступила в редколлегию 15.09.2012

УДК 621.396

А.А. СМИРНОВ, канд. техн. наук

КРИТЕРИИ И ПОКАЗАТЕЛИ ЭФФЕКТИВНОСТИ СТЕГАНОГРАФИЧЕСКИХ СИСТЕМ ЗАЩИТЫ ИНФОРМАЦИИ

Введение

Современное развитие информационных технологий привело к появлению новых угроз информационной безопасности, реализация которых может привести к непоправимым последствиям в различных сферах человеческой деятельности [1 – 4]. Следовательно, разработка перспективных технологий защиты информации, позволяющих противостоять современным угрозам, является актуальным научно-техническим заданием.

Одним из перспективных направлений в развитии современной теории защиты информации является цифровая стеганография, основной задачей которой является сокрытие не только смыслового содержания передаваемых данных, но и самого факта осуществления скрытной передачи информации [5 – 8]. Реализованные таким образом механизмы безопасности потенциально позволят обеспечить защиту от большего числа различных угроз информационной безопасности, предоставив конечному пользователю необходимые гарантии функций защиты.

В данной работе исследуется математическая модель и структурная схема стеганографической системы, вводятся критерии и показатели эффективности ее функционирования.

Структурная схема и математическая модель стеганографической системы

Основное функциональное отличие стеганографической и криптографической (секретной) систем определяется их целевым назначением. В то время как основной задачей секретной системы является сокрытие смыслового содержания передаваемых сообщений, основной задачей стеганографической системы является сокрытие самого факта организации передачи сообщений. Другими словами, противник, наблюдающий за открытым каналом связи в секретной системе, *наблюдает факт передачи данных*, и его задачей является поиск смыслового содержания этих данных. В стеганографической системе противник, наблюдая за открытым каналом связи, должен сперва *установить факт передачи данных*, после чего (при наличии такой передачи) найти смысловое содержание передаваемых данных.

Сокрытие факта организации передачи информационных сообщений безусловно обеспечивает и сокрытие их смыслового содержания. С этой точки зрения стеганографическую систему следует считать некоторым обобщением секретной системы, поскольку потеря (снижение) функции сокрытия факта организации передачи данных переводит стеганографическую систему в разряд секретной (криптографической) системы, ограничивая ее задачи до сокрытия смыслового содержания передаваемых сообщений.

Учитывая указанное функциональное отличие стеганографической системы, ее математическую опишем следующим образом.

Пусть $I = \{I_0, I_1, I_2, ..., I_m\}$ – множество возможных сообщений (событий), появляющихся на выходе источника информации, где символом I_0 формально обозначено «пустое сообщение», т.е. обозначение I_0 соответствует случаю, когда источник информации не формирует никаких сообщений, подлежащих передаче по каналам связи. Символы $I_1, I_2, ..., I_m$ соответствуют случаю формирования m различных «ненулевых» сообщений. Основной задачей секретной системы, как было показано выше, является сокрытие смыслового содержания этих информационных сообщений для любого $i = 1, 2, ..., m$. Сам факт передачи данных по каналам связи не скрывается и из формального описания в математической модели секретной системы случай отсутствия передачи «пустого сообщения» исключен.

Основной задачей стеганографической системы является такая организация передач сообщений, при которой для любого $I_i, i = 0, 1, 2, ..., m$ наблюдаемый противником открыт канал связи неразличим, т.е. передача любого из сообщений $I_1, I_2, ..., I_m$ для противни неотличима от передачи «пустого сообщения» I_0, когда информационные данные вов не передаются. Задача распознавания передачи «пустого сообщения» I_0 и передачи любо из «ненулевых» сообщений $I_1, I_2, ..., I_m$ и есть задачей установления факта передачи да ных, решаемой противником в стеганографической системе защиты информации.

Зафиксируем множество информационных данных $M = \{M_0, M_1, M_2, ..., M_m\}$, где ка дое $M_i, i = 1, 2, ..., m$ соответствует результату предварительного кодирования сообщений множества I, т.е. $M_i = f(I_i), i = 0, 1, 2, ..., m$, где $f(x)$ – формальное обозначение функци предварительного кодирования, реализуемой соответствующим устройством. Предварител ное кодирование предназначено для подготовки информационного сообщения к встраив нию в контейнер (например, посредством шифрования, помехоустойчивого кодировани и/или преобразования информационного сообщения в массив специально отформатирован ных цифровых данных). Символом M_0 формально обозначен результат предварительного кодирования «пустого сообщения», т.е. обозначение M_0 соответствует случаю отсутстви каких-либо информационных данных, подлежащих передачи по открытым каналам связи.

Зафиксируем множество возможных пустых контейнеров $L = \{L_1, L_2, ..., L_l\}$, предназна ченных для встраивания в них информационных данных. Обозначим чере $E^* = \{E_1^*, E_2^*, ..., E_n^*\}$ множество возможных стеганограмм, которое содержит все элемент множества пустых контейнеров L, а также все элементы множества заполненных контейне ров $E = \{E_1, E_2, ..., E_{lm}\}$. Другими словами, каждый элемент множества E^* однозначно свя зан с соответствующей парой элементов из множеств L и M соответственно. При этом множество E^* должно содержать все элементы множества L, поскольку в процессе стеганографического преобразования информационные данные могут вовсе не встраиваться в какой-либо пустой контейнер из множества L. Формально этот случай соответствует встраиванию «пустого сообщения» M_0, при этом соответствующие элементы множества E будут тождественны элементам множества L. Таким образом, $E = L \cup E^*$, причем $|E^*| = ml$, $|E^*| = n = (m+1)l$.

Зафиксируем множество отображений:

$$\varphi = \{\varphi_1, \varphi_2, ..., \varphi_k\},$$

где φ_i: $(M, L) \to E^*$, $i = 1, 2, ..., k$, а также множество обратных отображений:

$$\varphi^{-1} = \{\varphi_1^{-1}, \varphi_2^{-1}, ..., \varphi_k^{-1}\}$$

где каждое

$$\varphi_i^{-1}: E^* \to (M^*, L),$$

которое каждому элементу множества E^* ставит в соответствие элемент множества L и элемент множества M^*, причем

$$M^* = M \cup \overline{M},$$

где $\overline{M} = \{\overline{M_1}, \overline{M_2}, ..., \overline{M_l}\}$ – множество извлеченных данных из стеганограмм, соответствующих пустым контейнерам $L = \{L_1, L_2, ..., L_l\}$.

Другими словами, при реализации обратного стеганографического преобразования информации в результате извлечения имеем некоторую «оценку» информационных данных M_j^*, которая может как относиться, так и не относиться к множеству «ненулевых» информационных данных $\{M_1, M_2, ..., M_m\}$. Задачей уполномоченного пользователя на приемной стороне стеганографической системы является детектирование сообщения, т.е. сопоставление полученной «оценки» M_j^* с одним из подмножеств.

Зафиксируем множество ключей $K = \{K_1, K_2, ..., K_k\}$ так, что для всех $i = 1, 2, ..., k$ отображение $\varphi_i \in \varphi$ однозначно задается ключом K_i, т. е.:

$$\varphi_i : (M, L) \xrightarrow{K_i} E .$$

Каждое конкретное отображение φ_i из множества φ соответствует способу встраивания сообщения из множества M в контейнер из множества L при помощи конкретного ключа K_i. На рис. 1 схематично представлено отображение $\varphi_i \in \varphi$, заданное ключом K_i.

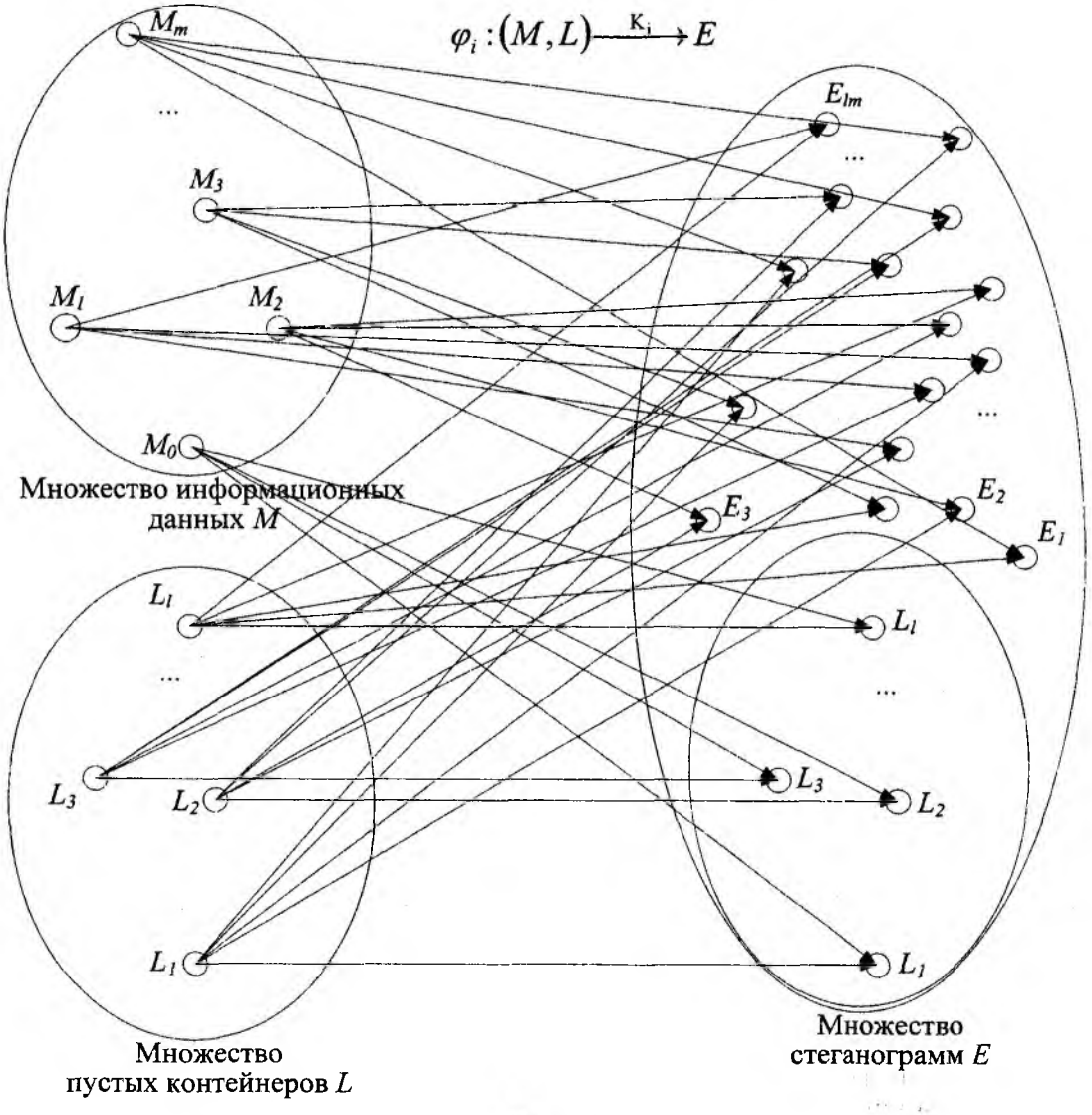


Рис. 1. Отображение $\varphi_i : (M, L) \xrightarrow{K_i} E$ множества открытых текстов и множества контейнеров в множество стеганограмм

Зафиксируем множество ключей $K^* = \{K_1^*, K_2^*, ..., K_k^*\}$, в общем случае $K \neq K^*$. Все элементы множества обратных отображений:

$$\varphi^{-1} = \{\varphi_1^{-1}, \varphi_2^{-1}, ..., \varphi_k^{-1}\}$$

задаются соответствующим ключом:

$$\varphi_i^{-1} : E \xrightarrow{K_i^*} (M^*, L).$$

Каждое конкретное отображение φ_i^{-1} из множества φ^{-1} соответствует способу извлечения сообщения из заполненного контейнера (и формирования пустого контейнера) при помощи ключа K_i^*. Если известен ключ K_i^*, то в результате выполнения операции извлечения возможен лишь единственный ответ – элемент некоторого множества M^* и элемент множества L:

$$(M_j^*, L_l) = \varphi_i^{-1}(E_w, K_i^*).$$

Множество M^* содержит все элементы множества M, кроме того, оно также содержит и другие элементы, соответствующие результату извлечения.

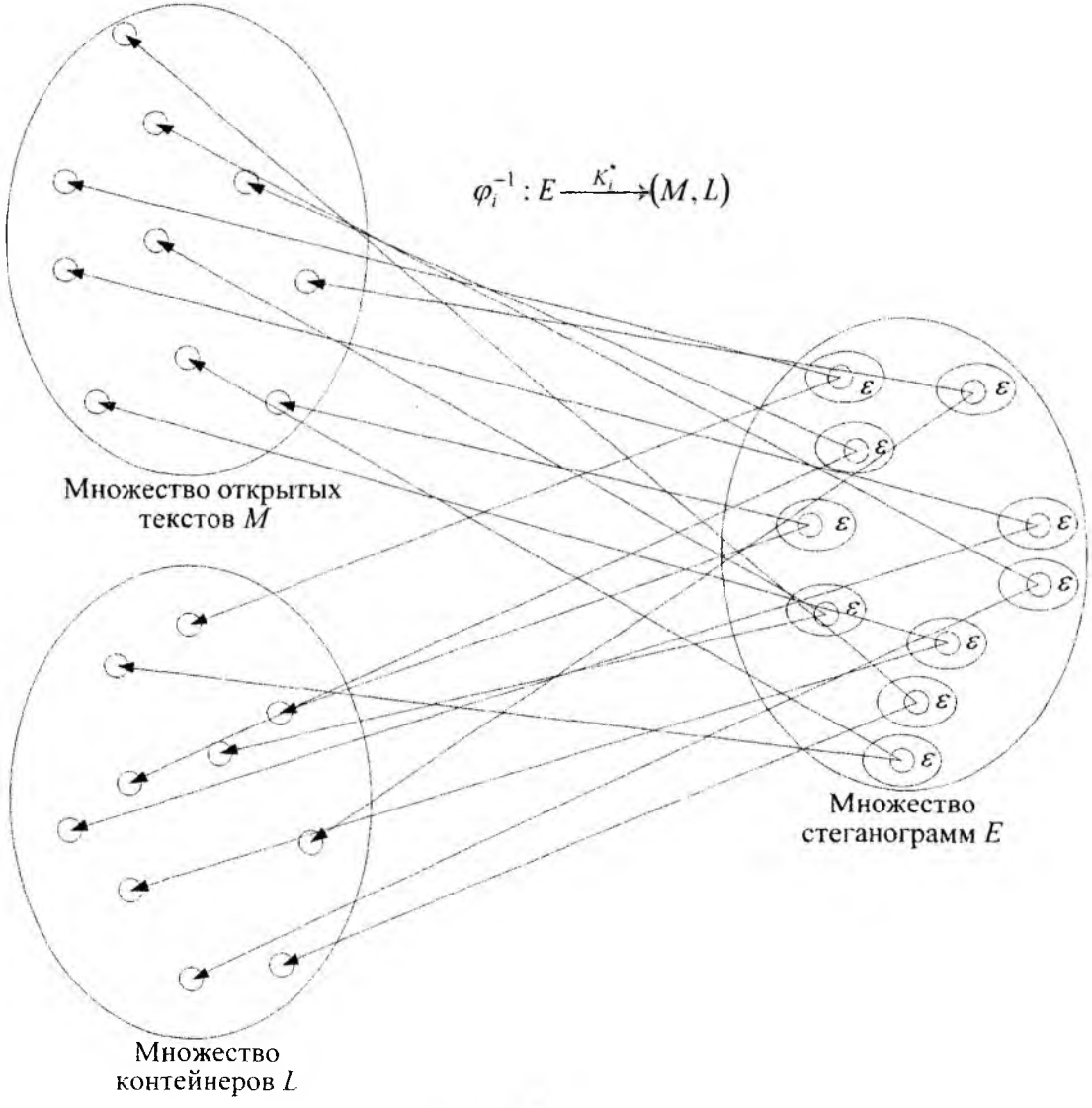


Рис. 2. Отображение $\varphi_i^{-1} : E \xrightarrow{K_i^*} (M, L)$ множества стеганограмм в множество открытых текстов и множество контейнеров

Для некоторых стеганографических систем справедливо равенство

$$\left(M_j, L_l\right) = \varphi_i^{-1}\left(E_w + \varepsilon, K_i^*\right),$$

т.е. незначительное изменение заполненного контейнера (на величину ε) не приведет к неправильному извлечению сообщения в процессе реализации отображения $\varphi_i^{-1} : E \xrightarrow{K_i^*} (M, L)$.

Этот процесс схематично представлен на рис. 2. Соответствующие указанному свойству стеганосистемы принято называть *робастными* [5 – 7].

Если для стеганосистемы характерно выполнение неравенства

$$\left(M_j, L_l\right) \neq \varphi_i^{-1}\left(E_w + \varepsilon, K_i^*\right)$$

для сколь угодно малой величины ε, тогда такие системы принято называть хрупкими.

Таким образом, в абстрактное *определение стеганографической системы* входят следующие множества M, L, E, φ, φ^{-1}, K и K^* (множества открытых текстов, пустых контейнеров и стеганограмм (заполненных контейнеров), множества прямых и обратных отображений и множества соответствующих им ключей).

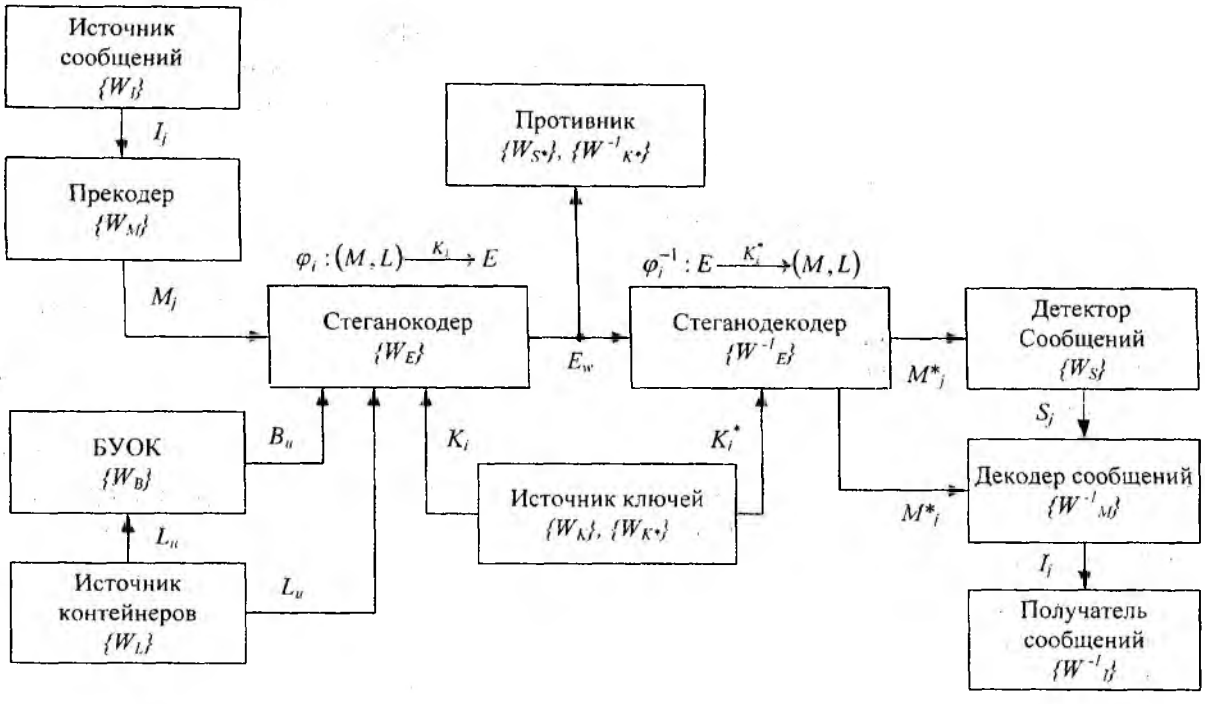


Рис. 3. Структурная схема стеганографической системы

На рис. 3 представлена структурная схема стеганографической системы. На схеме обозначены следующие элементы математической модели стеганосистемы:

– $I = \{I_1, I_2, \ldots, I_m\}$ – множество информационных сообщений, формируемых источником;

– $M = \{M_1, M_2, \ldots, M_m\}$ – множество информационных данных, полученных на выходе прекодера после соответствующего преобразования;

– $L = \{L_1, L_2, \ldots, L_l\}$ – множество пустых конейнеров, формируемых источником контейнеров и предназначенных для встраивания данных;

– $E = \{E_1, E_2, \ldots, E_n\}$ – множество возможных стеганограмм (заполненных контейнеров), формируемых устройством стеганографического кодирования (стеганокодером);

– $\varphi = \{\varphi_1, \varphi_2, \ldots, \varphi_k\}$ – множество прямых отображений множества информационных данных M и множества пустых контейнеров L в множество возможных стеганограмм E, формируемых стеганокодером, φ_i: $M \to E$, $i = 1, 2, \ldots, k$;

– $\varphi^{-1} = \{\varphi_1^{-1}, \varphi_2^{-1}, \ldots, \varphi_k^{-1}\}$ – множество обратных отображений множества возможных стеганограмм E в множество информационных данных M и множество пустых контейнеров L, $\varphi^{-1}{}_i$: $E \to M$, $i = 1, 2, \ldots, k$;

– $K = \{K_1, K_2, \ldots, K_k\}$ – множество ключей прямого стеганографического преобразования (стеганокодирования), причем каждое отображение $\varphi_i \in \varphi$ из множества прямых отображений $\varphi = \{\varphi_1, \varphi_2, \ldots, \varphi_k\}$ задается ключом K_i;

– $K^* = \{K_1^*, K_2^*, \ldots, K_k^*\}$ – множество ключей обратного стеганографического преобразования (стеганодекодирования), причем каждое обратное отображение $\varphi_i^{-1} \in \varphi^{-1}$ из множества обратных отображений $\varphi^{-1} = \{\varphi_1^{-1}, \varphi_2^{-1}, \ldots, \varphi_k^{-1}\}$ задается ключом K^*_i;

– $\{W_I\}$ – оператор формирования информационных сообщений $I_j \in \{M_1, M_2, \ldots, M_m\}$;

– $\{W_M\}$ – оператор предварительного кодирования информационных сообщений, результатом действия которого являются информационные данные $M_j \in \{M_1, M_2, \ldots, M_m\}$;

– $\{W_L\}$ – оператор формирования пустых контейнеров $L_u \in \{L_1, L_2, \ldots, L_l\}$, предназначенных для встраивания информационных данных;

– $\{W_B\}$ – оператор исследования и оценки особенностей формируемых контейнеров, результатом действия которого являются выработанные правила и ограничения B_u на встраивание информационных данных в сформированные контейнеры $L_u \in \{L_1, L_2, \ldots, L_l\}$;

– $\{W_E\}$ – оператор прямого стеганографического преобразования (стеганокодирования) информационных данных $M_i \in \{M_1, M_2, \ldots, M_m\}$ и пустых контейнеров $L_u \in \{L_1, L_2, \ldots, L_l\}$ в стеганограммы (заполненные контейнеры) $E_i \in \{E_1, E_2, \ldots, E_n\}$;

– $\{W^{-1}{}_E\}$ – оператор обратного стеганографического преобразования (стеганодекодирования) стеганограмм (заполненных контейнеров) $E_i \in \{E_1, E_2, \ldots, E_n\}$ в информационные данные $M_i \in \{M_1, M_2, \ldots, M_m\}$ и пустые контейнеры $L_u \in \{L_1, L_2, \ldots, L_l\}$;

– $\{W_K\}$ – оператор формирования ключевых данных $K_i \in \{K_1, K_2, \ldots, K_k\}$ прямого криптографического преобразования (шифрования) и ключевых данных $K^*_i \in \{K^*_1, K^*_2, \ldots, K^*_k\}$ обратного криптографического преобразования (расшифрования);

– $\{W^{-1}{}_{M^*}\}$ – оператор действий противника по реализации безключевого чтения, т.е. нахождения информационного сообщения $M_i \in \{M_1, M_2, \ldots, M_m\}$ без знания ключевых данных $K^*_i \in \{K^*_1, K^*_2, \ldots, K^*_k\}$ обратного криптографического преобразования (расшифрования);

– $\{W^{-1}{}_{K^*}\}$ – оператор действий противника по реализации поиска секретного ключа $K^*_i \in \{K^*_1, K^*_2, \ldots, K^*_k\}$ обратного криптографического преобразования (расшифрования);

– $\{W^{-1}{}_M\}$ – оператор обработки полученных информационных сообщений $M_i \in \{M_1, M_2, \ldots, M_m\}$ получателем информации.

Источник сообщений порождает поток информационных сообщений I_j из множества $I = \{I_1, I_2, \ldots, I_m\}$, которое, после предварительного преобразования в прекодере, формируется в виде сообщения M_j из множества M. Прекодер выполняет, таким образом, функцию предварительной подготовки информационного сообщения к встраиванию в контейнер (например, преобразование информационного сообщения в массив специально отформатированных цифровых данных и/или помехоустойчивое преобразование информации).

Источник контейнеров порождает поток пустых контейнеров L_u из множества $L = \{L_1, L_2, \ldots, L_l\}$. Сформированный контейнер L_u обрабатывается блоком учета особенностей контейнеров (БУОК). Основной функцией БУОК является выделение тех признаков (особенностей) B_u поступившего контейнера L_u, которые будут использованы при встраивании в него сообщения M_j.

Источник ключей в стеганографической системе порождает поток ключей из множества K и/или K^*. Выбор ключа K_i определяет конкретное отображение φ_i из множества отображе-

ний φ. С помощью отображения φ_i, соответствующего выбранному ключу K_i, по поступившему сообщению M_j и поступившему контейнеру L_u с учетом выявленных особенностей B_u контейнера L_u формируется стеганограмма (заполненный контейнер):

$$E_w = \varphi_i\left(K_i, M_j, L_u\right),$$

$$i \in [1,2,...,k],\ j \in [1,2,...,m],\ u \in [1,2,...,l],\ w \in [1,2,...,n],\ n \geq m.$$

Стеганограмма E_w передается в точку приема по некоторому каналу и может быть перехвачена противником. На приемном конце с помощью обратного отображения φ_i^{-1} (заданного ключом K_i^*) из стеганограммы E_w восстанавливается первоначальное сообщение и пустой контейнер:

$$\left(M_j, L_u\right) = \varphi_i^{-1}\left(K_i, E_w\right).$$

При передаче стеганограммы E_w по каналу связи и возможном воздействии противником на E_w передаваемая стеганограмма может исказиться. В этом случае на приемной стороне будет принята некоторая смесь переданного заполненного контейнера и результата воздействия на контейнер при передаче по каналу связи: $E_w+\varepsilon$. Выполнение операции обратного отображения φ_i^{-1} (заданного ключом K_i^*) в этом случае приведет к формированию некоторой оценки переданного сообщения и переданного пустого контейнера, т.е. получим:

$$\left(M_j^*, L_u^*\right) = \varphi_i^{-1}\left(K_i, E_w + \varepsilon\right).$$

Для хрупких стеганографических систем неравенство $M_j^* \neq M_j$ должно приводить к отбраковке сообщения, т.е. при малейшем искажении контейнера ($\varepsilon \neq 0$) извлеченная оценка M_j^* не должна приводить к прочтению встроенного сообщения (сообщение M_j разрушается при $\varepsilon \neq 0$).

Робастные стеганографические системы устойчивы к воздействию на заполненный контейнер. В введенных выше обозначениях это означает, что при $\varepsilon \neq 0$ извлеченная оценка M_j^* должна сопоставляться с одним из возможных сообщений (в идеальном случае, с сообщением M_j). В то же время, полученный из канала связи контейнер E_w может вовсе не содержать встроенного сообщения, т.е извлеченная из контейнера оценка M_j^* не должна быть сопоставлена ни с одним из допустимых сообщений. Функции детектирования встроенного сообщения на приемной стороне возложены на детектор сообщений, который по поступившей оценке M_j^* принимает решение о наличии или отсутствии встроенного сообщения в принятом контейнере E_w. Таким образом, оценка детектора S_j может быть интерпретирована как двоичное (да/нет) решение помехоустойчивого декодера о наличии или отсутствии неисправляемой ошибки. Само декодирование осуществляется в декодере сообщений, основными функциями которого является сопоставление извлеченной оценки M_j^* с одним из возможных сообщений M_j и преобразования последнего в информационное сообщение I_j, предоставляемое получателю информации.

Критерии и показатели эффективности стеганографических систем

Под эффективностью технической системы в широком смысле понимают соответствие результата выполнения некоторой операции требуемому [9]. При этом техническая система выступает в роли средства реализации исследуемой операции [10].

Применительно к рассматриваемому процессу стеганографическая система выступает в роли технического средства реализации операции, целью которой является сокрытие от противника факта осуществления скрытной передачи информации.

Таким образом, с учетом функционального назначения стеганосистемы, введем следующие показатели эффективности:

1. Пропускная способность – как предел отношения объема V встраиваемой в контейнер информации к общему объему D контейнера:

$$Q = \lim_{D\to\infty}\left(\frac{V}{D}\right). \tag{1}$$

2. Объем ключевых данных (в битах):

$$l_K = \log_2(|K|), \tag{2}$$

где $|K|$ – мощность множества ключевых данных.

3. Стойкость стеганосистемы.

3.1. Вероятность нахождения противником ключа детектирования (извлечения) сообщений, оцениваемая по критерию минимального риска:

$$P_K = \max\{P_K(u_1), P_K(u_2),..., P_K(u_{L_K}),\},$$

где $P_K(u_i)$ – вероятность нахождения противником ключа детектирования (извлечения) сообщений при использовании им стратегии u_i, L_K – количество различных стратегий противника по нахождению ключа детектирования (извлечения) сообщений.

3.2. Вероятность несанкционированного детектирования (извлечения) сообщения без знания секретного ключа (вероятность безключевого детектирования (чтения) сообщений), оцениваемая по критерию минимального риска:

$$P_M = \max\{P_M(v_1), P_M(v_2),..., P_M(v_{L_M}),\},$$

где $P_M(v_i)$ – вероятность несанкционированного детектирования (извлечения) сообщения противником при использовании им стратегии v_i, L_M – количество различных стратегий противника по несанкционированногому детектированию (извлечению) сообщения.

3.3. Безопасное время работы стеганосистемы, характеризующее вычислительные возможности противника по реализации различных стратегий (атак), оцениваемое по критерию минимального риска:

$$T_Б = \min_{i,j}\{T_Б(u_i), T_Б(v_j)\}, \tag{3}$$

где

$$T_Б(u_i) = \frac{(P_K(u_i))^{-1}}{\Psi\gamma},\ T_Б(v_i) = \frac{(P_K(v_j))^{-1}}{\Psi\gamma},$$

$T_Б(u_i)$, $T_Б(v_i)$ – безопасное время работы стеганосистемы, характеризующее вычислительные возможности противника по реализации различных стратегии (атаки) u_i и v_j, $i = 1, L_K$, $j = 1, L_M$, Ψ – производительность вычислительной системы, доступная противнику, измеряемая в количестве переборных операций в единицу времени, γ – числовой показатель (константа), для пересчета в требуемые единицы измерения безопасного времени (например, при оценке безопасного времени в годах соответствующее значение $\gamma \approx 3{,}2\cdot 10^7$).

4. Величина вносимых искажений как предел процентного отношение среднеарифметического всех абсолютных значений Δ-изменений данных контейнера к максимально возможному значению $\Delta_{\max}$:

$$I^* = \lim_{D\to\infty}\left(\frac{\Delta_{cp}}{\Delta_{\max}}\cdot 100\right) = \lim_{D\to\infty}\left(\frac{100}{\Delta_{\max}\cdot D}\cdot\sum_{i=1}^{D}|\Delta_i|\right), \quad (4)$$

де $\Delta_i - \Delta$ – изменения i-го элемента контейнера.

5. Вероятность ошибочного извлечения информационных данных сообщения как предел отношения числа ошибочно извлеченных сообщений:

$$P_{ош} = \lim_{V\to\infty}\frac{V_{ош}}{V} = 1 - \lim_{V\to\infty}\left(\frac{V - V_{ош}}{V}\right), \quad (5)$$

где $V_{ош}$ – объем ошибочно извлеченных данных.

Множество введенных показателей (1) – (5) и критериев их оценки позволяет математически формализовать постановку задачу совершенствования стеганосистемы.

Так, обобщенный показатель эффективности стеганографической системы защиты информации с учетом (1) – (5) в общем виде запишем как функционал:

$$W = F(Q, l_K, P_K, P_M, T_Б, I^*, P_{ош}),$$

где вид функционала $F(Q, l_K, P_K, P_M, T_Б, I^*, P_{ош})$ и конкретный вклад отдельных (частных) показателей эффективности $Q, l_K, P_K, P_M, T_Б, I^*, P_{ош}$ в значении обобщенного показателя W определяется исходя из конкретного назначения системы, особенностей ее использования и условий эксплуатации.

Целевую функцию задачи совершенствования стеганосистемы в общем виде запишем через максимизацию обобщенного показателя эффективности:

$$\max(W) = \max(F(Q, l_K, P_K, P_M, T_Б, I^*, P_{ош})).$$

Выводы

В ходе исследований введены основные элементы и математические операторы, абстрактно описывающие стеганографическую систему защиты информации. Во введенной формализации получено определение хрупких и робастных стеганосистем, а также введены критерии и показатели эффективности стеганографической защиты информации. В общем виде формализована задача совершенствования стеганосистемы через максимизацию обобщенного показателя ее эффективности. Перспективным направлением дальнейших исследований является сравнительный анализ известных стеганосистем по введенным показателям и критериям оценки эффективности, обоснование перспективных путей их дальнейшего совершенствования.

Список литературы: 1. *Final* report of European project number IST-1999-12324, named New European Schemes for Signatures, Integrity, and Encryption, April 19, 2004 – Version 0.15 (beta), Springer-Verlag, p 829. 2. *Alfred J. Menezes, Paul C. van Oorschot, Scott A. Vanstone.* Handbook of Applied Cryptography – CRC Press, 1997. – 794 p. 3. *Молдовян, Н.А., Молдовян, А.А., Еремеев, М.А.* Криптография: от примитивов к синтезу алгоритмов. – СПб. : БХВ-Петербург, 2004. – 448с. 4. *Сидельников, В.М.* Криптография и теория кодирования // Материалы конференции «Московский университет и развитие криптографии в России». – МГУ, 2002. – 22 с. 5. *Конахович, Г. Ф., Пузыренко, А. Ю.* Компьютерная стеганография. Теория и практика. – К. : «МК-Пресс», 2006. – 288 с. 6. *Грибунин, В. Г., Оков, И. Н., Туринцев, И. В.* Цифровая стеганография. Серия: Аспекты защиты. – Солон-Пресс, 2002. – 272 с. 7. *Хорошко, В.А., Шелест, М.Е.* Введение в компьютерную стеганографию. – К. : НАУ, 2002. – 140 с. 8. *Хорошко,В.О., Азаров,О.Д., Шелест, М.Е., Яремчук, Ю.Е.* Основи комп'ютерної стеганографії : навч. посібник. – Вінниця : ВДТУ, 2003. – 143 с. 9. *Авдуевский, В.С.* Надежность и эффективность в технике : справочник : в 10 т. / Ред. совет: В.С. Авдуевский (пред.) и др. – М. : Машиностроение, 1986. – Т. 1 : Методология. Организация. Терминология. – 224 с. 10. *Авдуевский, В.С.* Надежность и эффективность в технике : справочник : в 10 т. / Ред. совет: В.С. Авдуевский (пред.) и др. – М.: Машиностроение, 1986. – Т. 3. Эффективность технических систем. – 328 с.

Кировоградский национальный технический университет

Поступила в редколлегию 15.09.2012

УДК 621.391.7:336.71(075.8)

И.В. ОЛЕШКО, И.Д. ГОРБЕНКО, д-р техн. наук

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ПРОТОКОЛОВ СТРОГОЙ АУТЕНТИФИКАЦИИ

Введение

Аутентификация сообщений представляет собой процедуру, обеспечивающую связывающимся сторонам возможность проверки аутентичности (подлинности) получаемых сообщений. Двумя важными задачами аутентификации является проверка того, что содержимое сообщения не было изменено, и того, что сообщение прибыло именно из того источника, о котором информирует сообщение [1 – 3]. Одной из распространенных схем аутентификации является простая аутентификация, которая основана на применении традиционных многоразовых паролей с одновременным согласованием средств их использования и обработки. Идея строгой аутентификации, реализуемая в криптографических протоколах, заключается в следующем. Проверяемая (доказывающая) сторона доказывает свою подлинность проверяющей стороне, демонстрируя знание какого-либо секрета, который, например, может быть предварительно распределен безопасным способом между сторонами аутентификационного обмена. Доказательство знания секрета осуществляется с помощью последовательности запросов и ответов с использованием криптографических методов и средств.

Существенным является тот факт, что доказывающая сторона демонстрирует только знание секрета, но сам секрет в ходе аутентификационного обмена не раскрывается. Это обеспечивается посредством ответов доказывающей стороны на различные запросы проверяющей стороны. При этом результирующий запрос зависит только от пользовательского секрета и начального запроса, который обычно представляет собой произвольно выбранное в начале протокола большое число.

В большинстве случаев строгая аутентификация заключается в том, что каждый пользователь аутентифицируется по признаку владения своим секретным ключом. Иначе говоря, пользователь имеет возможность определить, владеет ли его партнер по связи надлежащим секретным ключом и может ли он использовать этот ключ для подтверждения того, что он действительно является подлинным партнером по информационному обмену.

В соответствии с рекомендациями стандарта X.509 различают процедуры строгой аутентификации следующих типов [4]:

- односторонняя аутентификация;
- двусторонняя аутентификация;
- трехсторонняя аутентификация.

Односторонняя аутентификация предусматривает обмен информацией только в одном направлении. Данный тип аутентификации позволяет:

1. Подтвердить подлинность только одной стороны информационного обмена;
2. Обнаружить нарушение целостности передаваемой информации;
3. Обнаружить проведение атаки типа «повтор передачи»;
4. Гарантировать, что передаваемыми аутентификационными данными может воспользоваться только проверяющая сторона.

Двусторонняя аутентификация по сравнению с односторонней содержит дополнительный ответ проверяющей стороны доказывающей стороне, который должен убедить ее, что связь устанавливается именно с тем партнером, которому были предназначены аутентификационные данные. Трехсторонняя аутентификация содержит дополнительную передачу данных от доказывающей стороны проверяющей. Такой подход позволяет отказаться от использования меток времени при проведении процедуры аутентификации. В зависимости от используемых криптографических алгоритмов протоколы строгой аутентификации можно разделить на следующие группы [2, 3]:

- на симметричных алгоритмах шифрования;

• на алгоритмах электронной цифровой подписи;

• на использовании криптографического контрольного значения;

• основанные на нулевых знаниях;

• на сертификатах с использованием преобразований в группах точек эллиптической кривой.

Введем следующие обозначения:

TokenAB – маркер, который отсылается предъявителем А проверяющему В;

eK(Z) – результат зашифрования данных *Z* по алгоритму симметричного шифрования, используя ключ К;

T_X – метка времени, созданная объектом X;

N_X – порядковый номер, созданный объектом X;

A – идентификатор объекта *A;*

B – идентификатор объекта *B;*

$sS_x(Z)$ – электронная цифровая подпись, полученная в результате преобразования над данными *Z*, используя личный ключ подписи S_x;

R_x – случайное число, созданное объектом x.

Цель статьи – сравнительный анализ основних протоколов строгой аутентификации на основании условных и безусловных критериев.

Строгая аутентификация, основанная на симметричных алгоритмах

Для работы протоколов аутентификации, построенных на основе симметричных алгоритмов, необходимо, чтобы проверяющий и доказывающий с самого начала имели один и тот же секретный ключ. Для закрытых систем с небольшим количеством пользователей каждая пара пользователей может заранее разделить его между собой. В больших распределенных системах, применяющих технологию симметричного шифрования, часто используются протоколы аутентификации с участием доверенного сервера, с которым каждая сторона разделяет знание ключа. Такой сервер распределяет сеансовые ключи для каждой пары пользователей всякий раз, когда один из них запрашивает аутентификацию другого.

Требования, предъявляемые к механизмам аутентификации, приведены в стандарте [5]. Все эти требования должны быть выполнены, в противном случае появляется возможность угрозы компрометации процесса аутентификации или даже невозможность выполнения аутентификации.

Ниже приводятся два примера протоколов аутентификации, специфицированных в ISO/IEC 9798-2 [5]. Эти протоколы предполагают предварительное распределение разделяемых секретных ключей.

Рассмотрим следующие варианты аутентификации:

1. Односторонняя аутентификация с одним проходом;

2. Двусторонняя аутентификация с тремя проходами.

Односторонняя аутентификация с одним проходом. В этом протоколе аутентификации предъявитель А инициирует процесс, а объект В проверяет его аутентичность. Уникальность и своевременность обеспечивается путем генерации и проверки метки времени и порядкового номера.

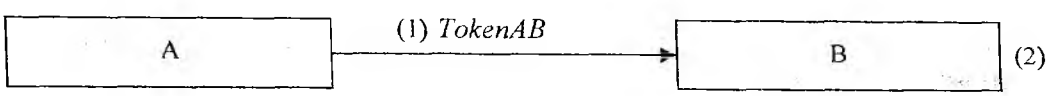


Рис. 1. Протокол односторонней аутентификации с одним проходом

Маркер (*TokenAB*), который отсылается предъявителем А проверяющему В, имеет такую структуру:

$$TokenAB = Text2 \,||\, eK_{AB}\begin{pmatrix} T_A \\ N_A \end{pmatrix} || \, B \,||\, Text1\Big). \qquad (1)$$

В маркере предъявитель А в качестве изменяемого во времени параметра использует или порядковый номер N_A, или метку времени T_A. Присутствие идентификатора В в маркере *TokenAB* не обязательно, если отсутствует возможность осуществления атаки типа маскарад.

Протокол осуществляется следующим образом:

1) Объект А генерирует и отправляет *TokenAB* объекту В.

2) После получения сообщения, которое содержит *TokenAB*, объект В проверяет *TokenAB* путем расшифрования зашифрованной части. Затем проверяется корректность идентификатора В, если он есть, а также метки времени или порядкового номера.

Двусторонняя аутентификация с тремя проходами. В этом протоколе аутентификации уникальность и / или своевременность обеспечиваются путем генерации и проверки случайного числа R_B. Протокол представлен на рис. 2.

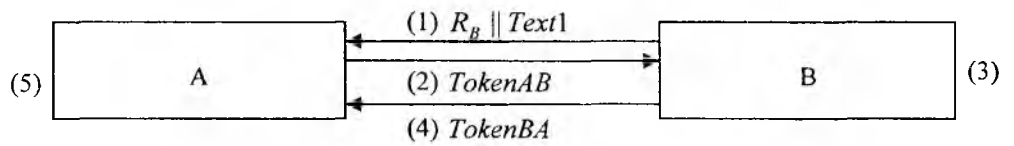


Рис. 2. Протокол двусторонней аутентификации с тремя проходами

Маркеры имеют такую структуру:

$$TokenAB = Text3 \,||\, eK_{AB}(R_A \,||\, R_B \,||\, B \,||\, Text2) \qquad (2)$$

$$TokenBA = Text5 \,||\, eK_{AB}(R_B \,||\, R_A \,||\, Text4). \qquad (3)$$

Протокол осуществляется следующим образом:

1) Объект В генерирует случайное число R_B и отправляет его объекту А.

2) Объект А генерирует случайное число R_A, формирует и отправляет объекту В маркер *TokenAB*

3) Объект В, получив сообщение, содержащее *TokenAB*, проверяет этот маркер путем расшифрования его зашифрованной части. Затем проверяет:

- корректность идентификатора В, если он есть.
- соответствие случайного числа R_B и случайного числа, которое содержится в маркере *TokenAB*.

4) Объект В генерирует и отправляет маркер *TokenBA* объекту А

5) Объект А, получив сообщение, содержащее *TokenBA*, проверяет этот маркер путем расшифрования его зашифрованной части. Затем проверяет: корректность идентификатора А, если он есть, случайного числа R_B и случайного числа, которое содержится в маркере *TokenBA*, соответствие случайного числа R_A, которое было отослано объекту В на шаге 2 и случайного числа, которое содержится в маркере *TokenBA*.

Протоколы строгой аутентификации, основанные на использовании электронной цифровой подписи

Рассмотрим следующие варианты протоколов аутентификации [6]:

1. односторонняя аутентификация с одним проходом;
2. двусторонняя аутентификация с тремя проходами.

Односторонняя аутентификация с одним проходом. Механизм изображен на рис. 3.

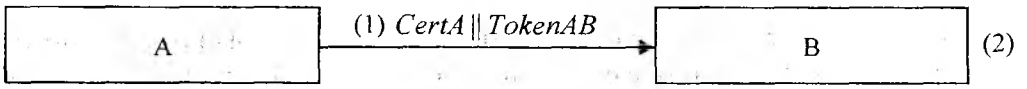


Рис. 3. Протокол односторонней аутентификации с одним проходом

Маркер (*TokenAB*), который отсылается предъявителем А проверяющему В, имеет такую структуру:

$$TokenAB = \begin{matrix} T_A \\ N_A \end{matrix} \| B \| Text2 \| sS_A\left(\begin{matrix} T_A \\ N_A \end{matrix} \| B \| Text1 \right) \quad (4)$$

Протокол осуществляется следующим образом:

1) Объект А генерирует и отправляет *TokenAB* объекту В, а также свой сертификат (не обязательно).

2) После получения сообщения, которое содержит *TokenAB*, объект В выполняет следующее:

- убеждается в том, что обладает действующим открытым ключом объекта А в виде сертификата или любых других средств;
- проверяет *TokenAB* путем:
- проверяет электронную цифровую подпись, которая содержится в маркере;
- проверяет корректность метки времени или порядкового номера;
- сравнивает идентичность значения поля идентификатора В в подписанных данных маркера *TokenAB* и идентификатора объекта В.

Двусторонняя аутентификация с тремя проходами. В этом протоколе аутентификации уникальность и/или своевременность обеспечивается путем генерации и проверки случайных чисел. Механизм изображен на рис.4.

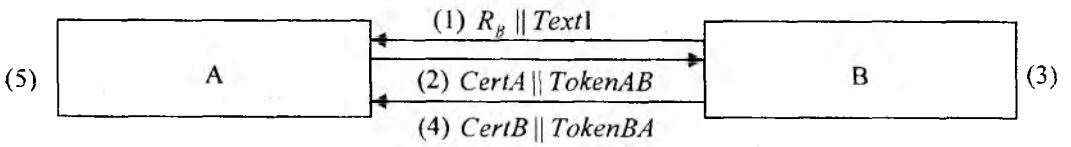


Рис. 4. Протокол двусторонней аутентификация с тремя проходами

Маркеры имеют следующую структуру:

$$TokenAB = R_A \| R_B \| B \| Text3 \| sS_A(R_A \| R_B \| B \| Text2) \quad (5)$$

$$TokenBA = R_B \| R_A \| A \| Text5 \| sS_B(R_B \| R_A \| A \| Text4) \quad (6)$$

Присутствие идентификатора В в маркере *TokenAB* и присутствие идентификатора А в маркере *TokenBA* необязательно и зависит от окружения, в котором осуществляется протокол аутентификации.

Протокол осуществляется следующим образом:

1) Объект В генерирует случайное число R_B и отправляет его объекту А (также может быть отправлено значение текстового поля *Text*1).

2) Объект А формирует и отправляет объекту В маркер *TokenAB*, а также свой сертификат (не обязательно).

3) После получения сообщения, которое содержит *TokenAB*, объект В выполняет следующие действия:

- убеждается в том, что обладает действующим открытым ключом объекта А в виде сертификата или любых других средств;
- проверяет *TokenAB* путем:

– проверяет электронную цифровую подпись, которая содержится в маркере;

– проверяет соответствие случайного числа R_B, которое было отослано объекту А на шаге (1) и случайного числа, которое содержится в подписанных данных маркера *TokenAB*;

– сравнивает идентичность значения поля идентификатора В в подписанных данных маркера *TokenAB* и идентификатора объекта В.

4) Объект В генерирует и отправляет объекту А маркер *TokenBA*, а также свой сертификат (необязательно).

5) После получения сообщения, которое содержит *TokenBA*, объект А выполняет действия, аналогичные тем, которые выполнял объект В на 3 шаге. Дополнительно происходит проверка соответствия случайного числа R_B, которое содержится в маркере *TokenBA* и случайного числа, полученного на шаге 1.

Протоколы строгой аутентификации, основанные на использовании криптографического контрольного значения

Рассматриваются следующие варианты протоколов [7]:

- односторонняя аутентификация с одним проходом;
- двусторонняя аутентификация с тремя проходами.

Односторонняя аутентификация с одним проходом. Механизм изображен на рис.1.

Маркер (*TokenAB*), который отсылается предъявителем А проверяющему В, имеет такую структуру:

$$TokenAB = \begin{matrix} T_A \\ N_A \end{matrix} \| Text2 \| f_{K_{AB}} \left(\begin{matrix} T_A \\ N_A \end{matrix} \| B \| Text1 \right) \quad (7)$$

В данном протоколе в качестве изменяемого во времени параметра используется или порядковый номер N_A, или метка времени T_A. Как определяется в ISO/ IEC 9798-1, $f_k(X)$ - криптографическое контрольное значение, где К – секретный ключ, Х – случайная строка данных. Протокол осуществляется следующим образом:

1. Объект А генерирует и отправляет *TokenAB* объекту В.

2. После получения сообщения, которое содержит *TokenAB*, объект В проверяет *TokenAB* путем проверки метки времени или порядкового номера, вычисляя $f_{K_{AB}} \left(\begin{matrix} T_A \\ N_A \end{matrix} \| B \| Text1 \right)$ и сравнивая полученное значение с криптографическим контрольным значением, которое содержится в маркере. Вместе с этим проверяется корректность идентификатора В (если он есть).

Двусторонняя аутентификация с тремя проходами. Механизм изображен на рис.2 Маркеры имеют следующую структуру:

$$TokenAB = R_A \| Text3 \| f_{K_{AB}} (R_A \| R_B \| B \| Text2) \quad (8)$$

$$TokenBA = Text5 \| f_{K_{AB}} (R_B \| R_A \| Text4) \quad (9)$$

Протокол осуществляется следующим образом:

1) Объект В генерирует случайное число R_B и отправляет его объекту А (также может быть отослано текстовое поле *Text1*).

2) Объект А генерирует случайное число R_A, формирует и отправляет объекту В маркер *TokenAB*

3) Объект В, получив сообщение, содержащее *TokenAB*, проверяет этот маркер, вычисляя значение $f_{K_{AB}} (R_A \| R_B \| B \| Text2)$ и сравнивая полученное значение с криптографическим контрольным значением, которое содержится в маркере. Таким образом, проверяется корректность идентификатора В (если он есть), а также соответствие случайного числа R_B, которое было отослано объекту А на шаге (1) и случайного числа, которое было использовано для генерации маркера *TokenAB*.

4) Объект В генерирует и отправляет маркер *TokenBA* объекту А.

5) Объект А, получив сообщение, которое содержит маркер *TokenBA*, проверяет его, вычисляя значение $f_{K_{AB}}(R_B \| R_A \| Text4)$ и сравнивая это значение с криптографическим контрольным значением, которое содержится в маркере. Таким образом, проверяется, что случайное число R_B, которое было получено от объекта В на шаге (1), и случайное число R_A, отосланное объекту В на шаге (2), были использованы для генерации маркера *TokenBA*.

Протоколы строгой аутентификации, которые используют методы, основанные на нулевых знаниях

Требования, выдвигаемые к данной группе протоколов, а также обоснование выбора основных параметров приведены в стандарте [8].

Рассмотрим следующие варианты протоколов:

- механизмы, основанные на использовании данных идентификации;
- механизмы, основанные на сертификатах с использованием системы асимметричного шифрования;
- механизмы, основанные на сертификатах с использованием дискретных логарифмов.

Механизмы, основанные на использовании данных идентификации. При выполнении однонаправленного механизма аутентификации предъявитель А инициирует процесс, а объект В проверяет его аутентичность. Для корректного выполнения механизма важно обеспечить объекта В данными идентификации объекта А, которые добавляются к одному из информационных обменов этого протокола. Одна итерация процедуры аутентификации приведена на рис.5.

Форма первого маркера ($TokenAB_1$) выглядит следующим образом:

$$TokenAB_1 = h(W \| Text) \tag{10}$$

где W – это доказательство, h – функция хеширования, а $Text$ – необязательное текстовое поле. Форма второго маркера имеет следующий вид:

$$TokenAB_2 = D \tag{11}$$

где D – ответ.

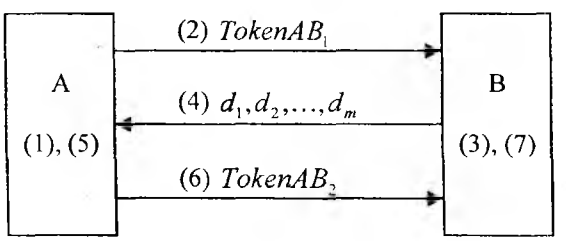


Рис. 5. Механизмы, основанные на использовании данных идентификации

Для каждого применения данного механизма подобная процедура аутентификации должна выполняться t раз. Проверяющий В должен считать, что предъявитель А прошел аутентификацию только при условии, что все t итераций завершились успешно. Протокол выполняется следующим образом:

1) Объект А, который владеет личной информацией аккредитации C_{A1}, C_{A2},..., C_{Am}, выбирает случайное число r, которое должно быть целым и удовлетворять неравенству $1 \le r \le n-1$. Это число сохраняется в тайне объектом А. Объект А вычисляет доказательство W:

$$W = r^v \ mod^* \ n \tag{12}$$

2) Объект А отправляет $TokenAB_1$ объекту В.

3) Получив маркер $TokenAB_1$, объект В должен случайным образом выбрать последовательность целых чисел $d_1,d_2,\ldots d_m$, где $0<d_i< \nu-1$.

4) Объект В отправляет запрос $d_1,d_2,\ldots d_m$ объекту А.

5) Получив запрос $d_1,d_2,\ldots d_m$, объект А должен посчитать ответ D из секретного значения r и личной информации аккредитации $C_{A1},C_{A2},\ldots,C_{Am}$ по следующей формуле:

$$D = r\prod_{i=1}^{m}(C_{Ai})^{d_i} \bmod^* n \qquad (13)$$

6) Объект А отправляет маркер $TokenAB_2$ объекту В.

7) Получив ответ D, объект В должен выполнить следующие вычисления:

- проверяет, что $0<D<\frac{n}{2}$. Если это не так, объект В не аутентифицирует объект А.
- вычисляет избыточные данные идентификации J_{A1}, J_{A2}, ..., J_{Am}, из данных идентификации I_{A1}, I_{A2},... I_{Am}
- вычисляет значение W' по формуле:

$$W' = D^{\nu}\prod_{i=1}^{m}(J_{Ai})^{d_i} \; mod^* n \qquad (14)$$

если W было отослано при первом обмене процедуры, то объект В проверяет, что вычисленное значение W' = W. Если h(W||Text) было отослано при первом обмене протокола, то объект В проверяет, что h(W'||Text) = h(W||Text). Если проверка прошла успешно, то вся итерация завершена успешно. В других случаях В не аутентифицирует объект А.

Механизмы, основанные на сертификатах с использованием системы ассиметричного шифрования. Ниже приведены обмены, которые необходимо совершать при выполнении однонаправленного механизма аутентификации между предъявителем А и объектом В, который проверяет его аутентичность. Механизм аутентификации приведен на рис.6.

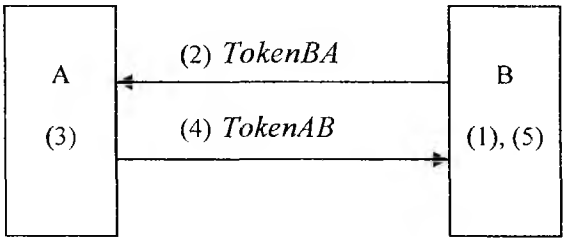


Рис. 6. Механизм, основанный на сертификатах
с использованием системы ассиметричного шифрования

Форма первого маркера (*TokenBA*), который отправляется проверяющим предъявителю

$$TokenBA = d \qquad (15)$$

где d – запрос.

Форма маркера (*TokenAB*), который отправляется предъявителем проверяющему выглядит следующим образом:

$$TokenAB = D \qquad (16)$$

где D – это ответ.

Протокол выполняется следующим образом:

1) Объект B выбирает случайное число r. Это число сохраняется в тайне объектом B. Затем объект B вычисляет h(r). Случайное число r должно быть выбрано таким способом, чтобы r||h(r) принадлежало к области P_A, открытого преобразования зашифрования объекта A. Объект B вычисляет запрос d по формуле

$$d = P_A(r \| h(r)) \tag{17}$$

2) Объект B отправляет *TokenBA* объекту A.

3) Получив маркер *TokenBA*, объект A должен выполнить следующее:

- объект A получает значение *r* путем вычисления

$$r \| h(r) = S_A(d) \tag{18}$$

где S_A – криптографическое преобразование расшифрования объекта A;

- объект A вычисляет значение $h(r)$ из полученного на шаге 3 значения r и сравнивает его со значением, полученным из маркера *TokenBA*. Если они не равны, A прекращает выполнение протокола, в другом случае A устанавливает значение ответа *D*: $D = r$.

4) Объект A отправляет $TokenAB = D$ объекту B.

5) Получив маркер *TokenAB*, объект B сравнивает значение ответа *D* и *r*. Если $r \neq D$, то считается, что механизм был провален, объект A не прошел аутентификацию. Если же $r = D$, то объект B признает аутентичность объекта A.

Механизмы, основанные на сертификатах с использованием дискретных логарифмов. Механизм аутентификации приведен на рис.7.

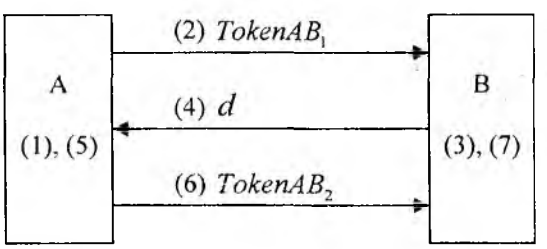


Рис. 7. Механизм, основанный на сертификатах с использованием дискретных логарифмов

Протокол выполняется следующим образом:

1) Объект A выбирает случайное число $1 \le r \le q$. Это число сохраняется в тайне объектом A. Объект A вычисляет доказательство W

$$W = g^r \bmod p \tag{19}$$

2) Объект A посылает $TokenAB_1$ объекту B. Маркер $TokenAB_1$ должен быть равен или W или h(W||Text);

3) Получив маркер $TokenAB_1$, объект B случайным образом выбирает $0 \le d \le q$;

4) Объект B отправляет запрос d объекту A;

5) Получив запрос d, объект A должен вычислить ответ D из секретного значения r и личного ключа z_A объекта A по формуле:

$$D = r - dz_A \bmod q \tag{20}$$

6) Объект A отправляет маркер $TokenAB_2$ объекту B.

7) Получив ответ D, объект B должен выполнить следующее:

- проверяет, что $0 < D < q$. И если это не так, то B отбраковывает A.
- объект B вычисляет W' по формуле

$$W' = (y_A)^d g^D \bmod p \quad (21,$$

Если W было отослано при первом обмене, то объект B проверяет, что значение W'=W. Если $h(W \| Text)$ было отослано при первом обмене, то B проверяет, значение $h(W' \| Text) = h(W \| Text)$.

Протоколы строгой аутентификации, основанные на сертификатах с использованием преобразований в группах точек эллиптической кривой

Рассмотрим согласование ключей типа Диффи-Гелмана (KANIDH) [9]. Этот механизм обеспечивает автономное установление разделяемого секрета между двумя объектами A и B.

Механизм протокола заключается в следующем:

1) Формирование ключа (A): объект A вычисляет разделяемый ключ, используя свой личный ключ для разделения ключей d_A и открытый ключ для разделения ключей P_A объекта A таким образом:

$$K_{AB} = (d_B \cdot l)(hP_A) \quad (22)$$

Свойства протокола:

- число проходов 0;
- механизм обеспечивает взаимную неявную аутентификацию ключа.

Критерии оценки криптографических протоколов

Криптографические протоколы будем оценивать на основании условных и безусловных критериев.

Безусловные критерии:

1. Надежность математической базы.
2. Практическая защищенность криптопреобразований от силовых и аналитических атак.
3. Реальная защищенность от всех известных и потенциально возможных криптоаналитических атак.
4. Статистическая безопасность криптографических преобразований.
5. Отсутствие слабых личных ключей.

Все обозначенные критерии не могут выполняться частично. Т.е. протокол может или удовлетворять поставленным условиям, или нет. При этом невыполнение хотя бы одного условия должно приводить к отказу от использования такого криптопреобразования, т.к. криптопреобразование является нестойким и его использование может привести к полному взлому системы криптографической защиты информации.

Условные критерии, используемые для сравнения криптографических протоколов:

- наличие и вид ключевой аутентификации. Обозначим этот критерий K_{y1};
- наличие и вид аутентификации субъекта. Обозначим этот критерий K_{y2};
- новизна ключей. Обозначим этот критерий K_{y3};
- управление ключами. Обозначим этот критерий K_{y4};
- эффективность протокола. Обозначим этот критерий K_{y5};
- криптоживучесть ключей. Обозначим этот критерий K_{y6};
- скорость выполнения элементарных операций и протокола в целом. Обозначим этот критерий K_{y7};

• уровень защищенности при реализации разных атак при разных условиях и отклонении свойств общесистемных параметров. Обозначим этот критерий $K_{уз}$.

Сравнительный анализ протоколов аутентификации

Сравнительный анализ протоколов аутентификации выполнен с использованием условных и безусловных критериев. Безусловные критерии выполняются для всех рассматриваемых протоколов. Анализ условных критериев приведен в таблице.

Таким образом, из таблицы можно сделать вывод о том, что более предпочтительным протоколом аутентификации является протокол, основанный на сертификатах с использованием преобразований в группе точек ЭК. За ним можно поставить протокол, основанный на использовании данных идентификации, затем протокол, основанный на сертификатах с использованием системы асимметричного шифрования и протокол, основанный на сертификатах с использованием дискретных логарифмов.

Протоколы / Критерии	Протокол, основанный на использовании данных идентификации	Протокол, основанный на сертификатах с использованием дискретных логарифмов	Протокол, основанный на сертификатах с использованием системы асимметричного шифрования	Протокол, основанный на сертификатах с использованием преобразований в группе точек ЭК
Аутентификация субъекта	Субъект А для субъекта В	Субъект А для субъекта В	Субъект А для субъекта В	Субъект А для субъекта В
Аутентификация ключа	Явная от А к У	Явная от А к У	Явная от А к У	Явная от А к У
Вид аутентификации субъекта	Односторонняя абонента А к У	Односторонняя абонента А к У	Односторонняя абонента А к У	Односторонняя абонента А к У
Вид аутентификации ключа	Односторонняя аутентификация ключа абонента А	Односторонняя аутентификация ключа абонента А	Односторонняя аутентификация ключа абонента А	Односторонняя аутентификация ключа абонента А
Наличие подтверждения ключа	Подтверждение ключа абонента А, C_A	Подтверждение ключа абонента А, Z_A	Подтверждение ключа абонента А, S_A	Подтверждение ключа абонента А, d_A
Новизна ключей	Абонент А владеет секретной информацией аккредитации C_A, а В знает только открытый ключ J_A C_A-статичный J_A- может меняться	Абонент А владеет секретным ключом Z_A, а В знает только открытый ключ Y_A, потому нет новизны ключей	Абонент А имеет собственное криптографическое преобразование S_A, а В знает только область открытого преобразования P_A, потому нет новизны ключей. P_A – определено однозначно, а S_A может меняться	Абонент А имеет секретный ключ d_A, а В знает только открытый ключ Y_A, потому нет новизны ключей, т.к. невозможно было бы вычислить значение W'
Управление ключевыми данными	Значение информации аккредитации C_A определяет доверенная сторона	По усмотрению сторон протокола	По усмотрению сторон протокола	По усмотрению сторон протокола

Защита от атаки типа «повтор ранее переданного сообщения»	Происходит за счет случайного числа r и последовательности целых чисел $d_1, d_2, d_3 K d_m$	Происходит за счет случайного числа r и случайного целого числа d	Происходит за счет случайного числа r	Происходит за счет случайного числа r и случайного целого числа d
Число обменов сообщениями	3	3	2	3
Сложность вычислений	1 операция секретного преобразования, 1 операция вычисления остаточных данных идентификации J_A и 1 операция открытого преобразования (3)	1 операция секретного преобразования, 1 операция открытого преобразования (2)	1 операция секретного преобразования, 1 операция открытого преобразования (2)	1 операция секретного преобразования 1 операция открытого преобразования (2)
Возможность использования предыдущих вычислений	Нет	Нет	Нет	Нет
Требования к 3-й стороне	Генерация для А информации аккредитации	Участники протокола сами решают, кто изготовляет пару ключей и если это третья сторона, то соответственно это генерация открытого и секретного ключа.	Участники протокола сами решают, кто изготовляет пару ключей и если это третья сторона, то соответственно это генерация открытого и секретного ключа.	Участники протокола сами решают, кто изготовляет пару ключей и если это третья сторона, то соответственно это генерация открытого и секретного ключа.
Криптостойкость ключа	Обеспечивается криптостойкость J_A ,при отсутствии компрометации секретной информации аккредитации C_A	Обеспечивается криптостойкость Y_A, при отсутствии компрометации секретной информации аккредитации Z_A	Обеспечивается криптостойкость P_A, при отсутствии компрометации секретной информации аккредитации S_A	Обеспечивается криптостойкость Y_A, при отсутствии компрометации секретной информации аккредитации d_A
Сложность реализации атак «полное раскрытие»	Субэкспоненциальная	Субэкспоненциальная	Субэкспоненциальная	Экспоненциальная
Неопровержимость	Неопровержимость объекта А происходит за счет C_A	Неопровержимость объекта А происходит за счет Z_A	Неопровержимость объекта А происходит за счет S_A	Неопровержимость объекта А происходит за счет d_A

Выводы

Расширение круга пользователей распределенных систем и усложнение задач, решаемых с помощью подобных сетей, привели к осознанию первостепенной важности проблем, связанных с обеспечением информационной безопасности при межсетевом взаимодействии. Поэтому одной из важных задач обеспечения безопасности информации является разработка

и анализ протоколов аутентификации, позволяющих одной стороне (проверяющему) убедиться в идентичности другой стороны (доказывающего). Одной из распространенных схем аутентификации является простая аутентификация, которая основана на применении традиционных многоразовых паролей с одновременным согласованием средств их использования и обработки. Одна из уязвимостей протоколов простой аутентификации заключается в том, что после того, как доказывающий передаст проверяющему свой пароль, проверяющий может выдать себя за проверяемого. Идея строгой аутентификации заключается в том, что проверяемая (доказывающая) сторона доказывает свою подлинность проверяющей стороне, демонстрируя знание какого-либо секрета, не раскрывая его.

В настоящее время на фоне огромного количества криптографических протоколов становится вопрос выбора того или иного протокола строгой аутентификации, который больше всего удовлетворяет поставленным требованиям. Выбрать такой протокол возможно только в случае полного анализа и сравнения характеристик криптографических протоколов.

В статье предлагается сравнивать криптографические протоколы аутентификации на основании условных и безусловных критериев. Основная задача безусловных критериев – выполнить анализ на предмет возможности применения выбранного криптографического протокола в принципе. Задача условных критериев – вывести числовые оценки для каждого криптографического протокола.

На основании анализа протоколов строгой аутентификации, был сделан вывод о том, что наилучшим протоколом аутентификации является протокол, основанный на сертификатах с использованием преобразований в группе точек ЭК.

Список литературы: 1. *Иванов, М. А.* Информации в компьютерных системах и сетях. М.: КУДИЦ-ОБРАЗ, 2001. – 368 с. 2. *Соколов, А.В.* Защита информации в распределенных корпоративных сетях и системах / А.В. Соколов, В.Ф. Шаньгин. – М. : ДМК Пресс, 2002. – 656 с. 3. *Шнайер, Б.* Прикладная криптография. Протоколы, алгоритмы, исходные тексты на языке Си / Б. Шнайер. – М. : ТРИУМФ, 2003. – 816 с. 4. *Х509-2001 – 4.* Використання каталогом основних положень сертифікації відкритого ключа та сертифікації атрибутів. 5. *ДСТУ* ISO/IEC 9798-2. Методи захисту. Автентифікація об'єктів. Частина 2: Механізми, що ґрунтуються на використанні алгоритмів симетричного шифрування. 6. *ДСТУ* ISO/IEC 9798-3. Методи захисту. Автентифікація об'єктів. Частина 3: Протоколи, що ґрунтуються на використанні електронного цифрового підпису. 7. *ДСТУ* ISO/IEC 9798-4. Методи захисту. Автентифікація об'єктів. Частина 4: Протоколи, що ґрунтуються на використанні функцій обчислення криптографічного контрольного значення. 8. *ISO/IEC* 9798-5. Методи захисту. Автентифікація об'єктів. Частина 5: Протоколи, що використовують методи які ґрунтуються на нульових знаннях. 9. *ISO/IEC* 15946-3:2002, IDT. Методи захисту криптографічні перетворення, що ґрунтуються на еліптичних кривих .

Харьковский национальный
университет радиоэлектроники

Поступила в редколлегию 20.09.2012

РАДИОЛОКАЦИЯ

УДК 621.396:621.391.82

О.В. СЫТНИК, д-р физ.-мат. наук, В.М. КАРТАШОВ, д-р техн. наук, А.А. СУПРУН

ПРОСТРАНСТВЕННАЯ СЕЛЕКЦИЯ ШИРОКОПОЛОСНЫХ ИСТОЧНИКОВ ПО СОБСТВЕННЫМ ЧИСЛАМ КОВАРИАЦИОННОЙ МАТРИЦЫ

Введение

Проблеме пространственно-временной обработки широкополосных сигналов в адаптивных антенных решетках (ААР) уделяется большое внимание при разработке радиолокационных систем различного назначения. Способность ААР оценивать угловые направления на источники сигналов с разрешением, существенно превышающим ограничения по критерию Рэлея [1], достигается благодаря специальным методам и алгоритмам обработки с использованием статистически достоверных оценок пространственно-временной ковариационной матрицы наблюдаемых сигналов [2]. Алгоритмы Эпплбаума [3], Уидроу [4], Прони [5], MUSIC [5], Кейпона [6], Гриффитса [7] и др., разработанные в 60-е – 80-е годы прошлого столетия, и ставшие уже классическими в технике пространственно-временной обработки сигналов в ААР. В качестве моделей сигналов используют, как правило, аппроксимацию узкополосных гауссовских процессов некоррелированных между собой и с собственным шумом приемников антенных элементов ААР.

При приеме широкополосных сигналов эффективность работы таких алгоритмов снижается. Для устранения этого недостатка в работе [2, 8] предложено в каждом канале ААР использовать многоотводную линию задержки, с помощью которой формируется пространственно-временная ковариационная матрица, и затем вычисляются весовые коэффициенты каждого канала ААР.

Цель настоящей работы – разработка статистически оптимальных алгоритмов пространственно-временной обработки сигналов ААР для получения эффективных оценок углового положения группы широкополосных источников сигналов.

Модель наблюдения

Рассмотрим ААР, состоящую из $M+1$ идентичных всенаправленных широкополосных антенных элементов, расположенных на одинаковых расстояниях d друг от друга. Сигнал с антенного элемента с индексом 0 поступает на сумматор, формирующий выходной сигнал ААР без усиления, т.е. весь спектр входных сигналов, передается на выход ААР без искажений. Выходные сигналы остальных M антенных элементов проходят через многоотводные линии задержки, усиливаются усилителями с комплексными коэффициентами усиления $\dot{w}_i$, и суммируются с сигналом антенного элемента с индексом 0.

В дальней зоне ААР находятся N источников ($N < M$), излучающих сигналы мощностью s_n, где $n = 1,2,...N$. s_0 – мощность собственного теплового шума приемной системы. Направление прихода фронта волны относительно нормали к азимутальной плоскости решетки каждого из источников обозначим θ_n, центральную частоту несущей – f_0, эффективную ширину спектра – β.

Спектральная плотность сигналов, поступающих на входы усилителей приемных каналов ААР от n-го источника, в общем виде может быть записана как

$$S_n(f) = \frac{s_n}{\beta \cdot f_0}\left[\dot{S}\left(\frac{2\Delta f}{\beta \cdot f_0}\right) + \dot{S}\left(\frac{-2\Delta f}{\beta \cdot f_0}\right)\right], \qquad (1$$

где $\Delta f = f - f_0$.

Применяя обратное преобразование Фурье и разложение в ряд Маклорена [9], представим ковариационную функцию в виде

$$R_n(t) = s_n \cdot \sum_{p=0}^{\infty} \alpha_{np} \frac{(\pi\beta f_0 t)^p}{p!} \cos\left(\omega_0 t + \frac{p\pi}{2}\right), \tag{2}$$

где

$$\alpha_{np} = \int_{-1}^{1} \dot{S}_n(f) f^p df, \; p = 0,1,2,\ldots,\infty \tag{3}$$

Очевидно, что каждому α_{np} можно поставить во взаимно однозначное соответствие p-й момент нормализованной спектральной плотности $\dot{S}_n(f)$, который пропорционален производной автоковариационной функции.

Обозначим вектор входных сигналов как $\vec{X}$

$$\vec{X} = [x_1, x_2, \ldots, x_{MJ}]^T, \tag{4}$$

где верхний индекс T обозначает операцию транспонирования.

Представление ковариационной матрицы

Соответствующая ковариационная матрица

$$\vec{R} = E\left\{\vec{X}\vec{X}^T\right\}, \tag{5}$$

где $E\{\cdot\}$ – символ операции вычисления математического ожидания.

С учетом (2) ковариационная матрица (5) может быть представлена в виде

$$\vec{R} = \sum_{p=0}^{\infty} \sum_{n=0}^{N} A_{np} \beta^n, \tag{6}$$

где нижний индекс np указывает на p-ю спектральную компоненту n-го источника сигнала.

Рассмотрим элемент (g,h) множества $\{A_{np}\}$, где g – соответствует j_g отводу от линии задержки m_g антенного элемента и h соответствует j_h отводу от линии задержки m_h антенного элемента, т.е. $g = m_g J + j_g$ и $h = m_h J + j_h$. Тогда с учетом (2) n-й источник формирует сигнал

$$\begin{aligned} \left|A_{np}\right|_{gh} &= \frac{s_n \alpha_{np} \pi^p}{p!} f_0^p \left[\left(j_g\tau + m_g\tau_n\right) - \left(j_h\tau + m_h\tau_n\right)\right]^p \times \\ &\times \cos\left[\omega_0\left(j_g\tau + m_g\tau_n\right) - \left(j_h\tau + m_h\tau_n\right) + \frac{p\pi}{2}\right], \end{aligned} \tag{7}$$

где

$$\tau_n = d\sin(\theta_n / c) \tag{8}$$

представляет собой задержку во времени прихода фронта плоской волны между соседними антенными элементами от n-го источника.

Для случая, когда шум приемника пренебрежимо мал, (7) можно записать как:

$$\left|A_{0p}\right|_{gh} = \begin{cases} 0, & m_g \neq m_h \\ \dfrac{s_0 \alpha_{0p} \pi^p}{p!}\left[f_0 \tau (j_g - j_h)\right]^p, & m_g = m_h. \end{cases}$$

Подставляя (7) и (9) в (6) после некоторых преобразований получим выражение матрицы $\vec{R}$

$$\vec{R} = \left[\beta^0 U_0 \ldots \beta^I U_I\right] \begin{bmatrix} E_{00} & \ldots & E_{0I} \\ \ldots & \ldots & \ldots \\ E_{I0} & \ldots & E_{II} \end{bmatrix} \begin{bmatrix} \beta^0 U_0^T \\ \vdots \\ \beta^I U_I^T \end{bmatrix}, \qquad (10$$

где $I = J/2 - 1$ и U_p, и E_{pq} имеют следующую структуру.

Набор матриц размера $(MJ \times 2M)$ со взаимно ортогональными столбцами $\vec{U}_p^T \vec{U}_q = 0$ $p \neq q$, обозначим $\vec{U}_0, \vec{U}_1, \ldots, \vec{U}_I$ полученных процедурой ортогонализации Грамма – Шмидт [10] матриц $\vec{C}_0, \vec{C}_1, \ldots, \vec{C}_I$ где

$$\begin{aligned} \vec{U}_0 &= \vec{C}_0 \\ \vec{U}_1 &= \left[\vec{I}_{MJ} - \vec{U}_0 \left(\vec{U}_0^T \vec{U}_0\right)^{-1} \vec{U}_0^T\right] \vec{C}_1 \\ \ldots &\quad \ldots \quad \ldots \\ \vec{U}_0 &= \left[\vec{I}_{MJ} - \sum_{p=0}^{I-1} \vec{U}_p \left(\vec{U}_p^T \vec{U}_p\right)^{-1} \vec{U}_p^T\right] \vec{C}_I, \end{aligned} \qquad (11$$

где $\vec{C}_p$ определяется как:

$$\vec{C}_p = (f_0 \tau)^p \begin{bmatrix} \cos(\omega_0 \tau + \dfrac{\pi p}{2}) & \ldots & -\sin(\omega_0 \tau + \dfrac{\pi p}{2}) \\ \vdots & \ldots & \vdots \\ J^p \cos(J\omega_0 \tau + \dfrac{\pi p}{2}) & \ldots & -J^p \sin(J\omega_0 \tau + \dfrac{\pi p}{2}) \end{bmatrix} \otimes \vec{I}_M, \qquad (12$$

где символ $\otimes$ – символ Кронекера; $\vec{I}_M$ – единичная матрица размера $(M \times M)$.

Поскольку $\vec{C}_p$ зависит только от величины задержки сигнала в элементе и от числ элементов, то и матрицы $\vec{U}_p$ также зависят только от этих параметров. В этом случа $(2M \times 2M)$ матрица сигналов $\vec{E}_{pq} = \vec{E}_{qp}^T$, а ее разложение в ряд Маклорена [9] можно пре ставить в виде

$$\vec{E}_{pq} = \frac{\pi^{p+q}}{p!q!} \sum_{n=0}^{N} s_n \alpha_{n,p+q} \vec{Q}_n \vec{Q}_n^T + 0(s), \qquad (13)$$

где $\vec{Q}_n^T = \vec{I}_{2M}$, если $n = 0$ и

$$\vec{Q}_n^T = \begin{matrix} \cos(\omega_0\tau_n)\sin(\omega_0\tau_n) & \dots & \cos(M\omega_0\tau_n)\sin(M\omega_0\tau_n) \\ -\sin(\omega_0\tau_n)\cos(\omega_0\tau_n) & \dots & -\sin(M\omega_0\tau_n)\cos(M\omega_0\tau_n) \end{matrix}, \text{если } n=1,\dots,N. \quad (14)$$

Используя разложение (13) представим ковариационную матрицу наблюдаемых сигналов в виде:

$$\vec{R}_{pq}^0 = \left(\vec{R}_{pq}^0\right)^T = \frac{\pi^{p+q}}{p!q!}\sum_{n=0}^{N} s_n \alpha_{n,p+q}\vec{Q}_n\vec{Q}_n^T . \quad (15)$$

Очевидно, что при использовании аппроксимации первого порядка, модель ковариационной матрица зависит только от источников сигналов и не зависит от величины задержки сигнала в элементе и от их числа, а также не зависит от эффективной ширины спектра сигнала. В этом случае полученная модель есть ни что иное как ковариационная матрица узкополосного сигнала для спектральной компоненты s_n взвешенная с множителем $\frac{\pi^{p+q}}{p!q!}\alpha_{n,p+q}$.

Учитывая (15), выражение для матрицы $\vec{R}$ можно представить в виде

$$\vec{R} = \begin{bmatrix} \beta^0\vec{U}_0 & \dots & \beta^l\vec{U}_l \end{bmatrix} \begin{bmatrix} R_{00}^0 & \dots & R_{0l}^0 \\ \vdots & \dots & \vdots \\ R_{l0}^0 & \dots & R_{ll}^0 \end{bmatrix} \begin{bmatrix} \beta^0\vec{U}_0^T \\ \vdots \\ \beta^l\vec{U}_l^T \end{bmatrix} . \quad (16)$$

Представление ковариационной матрицы в виде (17) интересно еще и тем, что она может быть разделена на две части соответствующие источникам $\vec{R}_{pq}^0$ и антенной решетки $\beta\vec{U}_p$. Покажем это на примере антенной решетки, состоящей из трех антенных элементов с четырьмя линиями задержки в приемных каналах. Величина задержки $\tau = 1/4f_0$. В дальней зоне такой антенны находится один источник сигнала, спектральная плотность сигнала которого равномерна во всей полосе. Таким образом, $M=2$, $J=4$, $f_0\tau = 1/4$, $N=1$ и $\dot{S}_1(f) = \dot{S}_0(f) = 1/2$ если $-1 < f < 1$ и $\dot{S}_0(f) = 0$ для всех других значений частоты f.

Согласно (3) $\alpha_{1p} = \alpha_{0p}$, а $\alpha_{00} = 1$, $\alpha_{01} = 0$ и $\alpha_{03} = 1/3$. Используя соотношения (10) – (12) и учитывая аппроксимацию первого порядка ковариационной матрицы (13), (14) получим выражение для матрицы $\vec{R}$ в виде (16)

$$\vec{R}_1 = \begin{bmatrix} 0 & 1 & 0 & \beta_{14} & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & -\beta_{14} & 0 & 0 \\ 0 & 1 & 0 & \beta_{14} & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -\beta_{14} & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & -\beta_{14} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\beta_{14} \\ 0 & 0 & 1 & 0 & 0 & \beta_{14} & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & -\beta_{14} \end{bmatrix} \times$$

$$\times \begin{bmatrix} s_0+s_1 & 0 & s_1\cos(\omega_0\tau_n) & s_1\sin(\omega_0\tau_n) & 0 & 0 & 0 & 0 \\ 0 & s_0+s_1 & -s_1\sin(\omega_0\tau_n) & s_1\cos(\omega_0\tau_n) & 0 & 0 & 0 & 0 \\ s_1\cos(\omega_0\tau_n) & -s_1\sin(\omega_0\tau_n) & s_0+s_1 & 0 & 0 & 0 & 0 & 0 \\ s_1\sin(\omega_0\tau_n) & -s_1\cos(\omega_0\tau_n) & 0 & s_0+s_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \pi^2(s_0+s_1)/3 & 0 & \pi^2 s_1\cos(\omega_0\tau_n)/3 & \pi^2 s_1\sin(\omega_0\tau_n)/3 \\ 0 & 0 & 0 & 0 & 0 & \pi^2(s_0+s_1)/3 & -\pi^2 s_1\sin(\omega_0\tau_n)/3 & \pi^2 s_1\cos(\omega_0\tau_n)/3 \\ 0 & 0 & 0 & 0 & \pi^2 s_1\cos(\omega_0\tau_n)/3 & -\pi^2 s_1\sin(\omega_0\tau_n)/3 & \pi^2(s_0+s_1)/3 & 0 \\ 0 & 0 & 0 & 0 & \pi^2 s_1\sin(\omega_0\tau_n)/3 & \pi^2 s_1\cos(\omega_0\tau_n)/3 & 0 & \pi^2(s_0+s_1)/3 \end{bmatrix} \times$$

$$\times\begin{bmatrix} 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ \beta_{14} & 0 & \beta_{14} & 0 & 0 & 0 & 0 & 0 \\ 0 & -\beta_{14} & 0 & -\beta_{14} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \beta_{14} & 0 & \beta_{14} & 0 \\ 0 & 0 & 0 & 0 & 0 & -\beta_{14} & 0 & -\beta_{14} \end{bmatrix} \quad (17)$$

Собственные числа ковариационной матрицы

Поскольку $\vec{R}^0_{pq} = \left(\vec{R}^0_{pq}\right)^T$ вторая матрица в (16) может быть представлена своим разложением как

$$\begin{bmatrix} R^0_{00} & \dots & R^0_{0l} \\ \vdots & \dots & \vdots \\ R^0_{l0} & \dots & R^0_{ll} \end{bmatrix} = \begin{bmatrix} R^0_{00} \\ \vdots \\ R^0_{l0} \end{bmatrix} \left(R^0_{00}\right)^{-1} \begin{bmatrix} R^0_{00} \\ \vdots \\ R^0_{l0} \end{bmatrix}^T + \begin{bmatrix} 0 \\ R^1_{1l} \\ \vdots \\ R^1_{ll} \end{bmatrix} \left(R^1_{ll}\right)^{-1} \begin{bmatrix} 0 \\ R^1_{1l} \\ \vdots \\ R^1_{ll} \end{bmatrix}^T + \dots + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ R^1_{ll} \end{bmatrix} \left(R^1_{ll}\right)^{-1} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ R^1_{ll} \end{bmatrix}^T, \quad (18)$$

где $\vec{R}^{r+1}_{pq} = \vec{R}^r_{pq} - \vec{R}^r_{pr}\left(\vec{R}^r_{rr}\right)^{-1}\left(\vec{R}^r_{qr}\right)^T$, $r = 0,\dots,l-1$, $p,q = r,\dots,l$. Матрица $\vec{R}^p_{pp}$ – положительно определенная несингулярная матрица.

Подставив (18) в (10) получим следующее представление матрицы $\vec{R}$:

$$\vec{R} = \begin{bmatrix} \vec{V}_0 & \dots & \vec{V}_l \end{bmatrix} \begin{bmatrix} \beta^0\left(\vec{R}^0_{00}\right)^{-1} & \dots & 0 \\ \vdots & \dots & \vdots \\ 0 & \dots & \beta^{J-2}\left(\vec{R}^l_{ll}\right)^{-1} \end{bmatrix} \begin{bmatrix} \vec{V}_0^T \\ \vdots \\ \vec{V}_l^T \end{bmatrix}, \quad (19)$$

где $\vec{V}_p = \beta^0 \vec{U}_p \vec{R}^p_{pp} + \dots + \beta^{l-p} \vec{U}_l \vec{R}^p_{lp}$.

Из последнего выражения следует, что поскольку столбцы $\vec{U}_0,\dots,\vec{U}_l$ взаимно ортогональны, то и столбцы $\vec{V}_0,\dots,\vec{V}_l$ также будут ортогональны. Это означает, что уравнение (19) является блочно-спектральным разложением ковариационной матрицы.

Анализ собственных чисел ковариационной матрицы свидетельствует, что даже если источники помеховых сигналов имеют равномерную спектральную плотность мощности в полосе приемной антенны, собственные числа матрицы можно разделить на два набора. Один из них соответствует сигналам источников помех, действующих на антенную решетку в области боковых лепестков, а другой, значения собственных чисел ковариационной матрицы в котором по абсолютной величине меньше чем в первом наборе, порожден тепловым собственным шумом приемников системы наблюдения

Выводы

Таким образом, так же как и в случае узкополосной адаптивной антенной решетки для широкополосных приемных систем селекция помеховых источников осуществляется на основе анализа собственных чисел пространственно-временной ковариационной матрицы сформированной с учетом наличия линий задержки в каждом из каналов приема. Блочная структура ковариационной матрицы, вычисляемой процессором ААР, эквивалентна разбиению всей полосы частот на несколько частотных участков. В каждом из которых вычисляются парные корреляционные связи между компонентами сигнала, принятыми разными антенными элементами в один и тот же момент времени. Этим обеспечивается сохранение инфор-

мации об углах падения плоских волн порожденных источниками помех на плоскость решетки. В последующем эти данные используются для вычисления вектора весовых коэффициентов адаптивной антенны. Очевидно, что количество элементов весового вектора будет равно произведению числа антенных элементов решетки на число линий задержки.

Список литературы 1. *Арманд, Н.А., Крапивин, В.Ф., Мкртчан, Ф.А.* Методы обработки данных радиофизического исследования окружающей среды. – М. : Наука, 1987. – 270 с. 2. *Монзинго, Р.А., Миллер, Т.У.* Адаптивные антенные решетки – М. : Радио и связь. – 1986. – 448 с. 3. *Applebaum, S.P.* Adaptive Arrays // IEEE Trans. Antennas Propag. – 1976. – Vol. AP-24, №5, September. – P. 585 – 598. 4. *Widrow, B., Mantey, P.E., Griffits, L.J. at all.* Adaptive Antenna Systems // Proc. IEEE. – 1967. – Vol. 55, December. – P. 2143 – 2159. 5. *Марпл, С.Л.* Цифровой спектральный анализ и его приложения. – М. : Мир, 1990. – 584 с. 6. *Capon, J.* Application of Detection and Estimation Theory to Large Array Seismology. // Proc. IEEE. – 1970. – Vol. 58, May. – P. 760 – 770. 7. *Griffiths, L.J.* A simple Algorithm for Real Time Processing in Antennas Arrays // Proc. IEEE. – 1969. – Vol. 57, October. – P. 1696 – 1707. 8. *Lacoss, R.T.* Adaptive Combining of Wideband Array Data for Optimal Reception // IEEE Trans. Geosci. Electron., – 1968. – Vol. GE-6, May. – P. 78 – 86. 9. *Корн, Г., Корн, Т.* Справочник по математике для научных работников и инженеров. – М. : Наука, 1984. – 832 с. 10. *Ланкастер, П.* Теория матриц : пер с англ. – М. Наука, 1982. – 272 с.

Харьковский национальный
университет радиоэлектроники

Поступила в редколлегию 17.09.2012

УДК 621.396

О.О. БАЙЗДРЕНКО, канд. техн. наук, В.Б. ЛУБСЬКИЙ

РОЗРАХУНОК ХАРАКТЕРИСТИК ВИЯВЛЕННЯ СИГНАЛІВ, ВІДБИТИХ ОБ'ЄКТАМИ, РОЗТАШОВАНИМИ НА МОРСЬКІЙ ПОВЕРХНІ, ПОЛУАКТИВНОЮ БІСТАТИЧНОЮ РЛС З ЦИФРОВИМ ТЕЛЕВІЗІЙНИМ СИГНАЛОМ ОСВІТЛЕННЯ

У 90-х роках XX ст. у Франції, Великій Британії, США, Білорусі і деяких інших країнах отримала практичний розвиток технологія полуактивної бістатичної радіолокації з зовнішнім освітленням від аналогових телевізійних передатчиків. Основні переваги таких радіолокаційних систем обумовлені відсутністю власного передатчика. При цьому досягається абсолютна радіотехнічна скритність, не потрібно призначення дефіцитного радіочастотного ресурсу, відсутній шкідливий вплив на екологію навколишнього середовища та на людину. Ще більше можливостей для полуактивної бістатичної радіолокації надає сигнал цифрового телевізійного мовлення, який суттєво переважає сигнал аналогового телебачення за наступними показниками: кращі енергетика, стабільність та потенційна роздільна здібність по часу затримки, нижчий рівень бокових пелюстків функції неоднозначності. Крім того, сигнал цифрового телебачення дозволяє здійснювати однозначну оцінку часу і частоти. Повний перехід України з 2015 року на цифрове телебачення стандарту DVB-T2 створює добрі перспективи для розвитку полуактивної бістатичної радіолокації з цифровим телевізійним освітленням.

Тема полуактивних бістатичних РЛС з телевізійним сигналом освітлення детально розроблена у роботах [1, 2, 17 – 21]. Усі вони присвячені дослідженню наземних систем виявлення повітряних цілей. У той же час, аналіз інфраструктури телемовлення у морських прибережних районах вказує, що берегова компонента систем освітлення обстановки на мор також має реальну можливість отримання додаткового каналу інформації за рахунок використання відбитих надводними об'єктами телевізійних сигналів у межах радіогоризонту. Оскільки наведені вище роботи не містять інформації стосовно застосування полуактивних бістатичних РЛС з телевізійним сигналом освітлення для виявлення надводних об'єктів, є актуальною розробка відповідних методик розрахунків характеристик виявлення, які враховують особливості обробки телевізійних сигналів, відбитих морськими цілями.

Метою даної статті є розробка варіанту розрахунку характеристик виявлення сигналів відбитих об'єктами, розташованими на морській поверхні, бістатичною радіолокаційною станцією з цифровим телевізійним сигналом освітлення.

Як вказувалося вище, одним з сигналів, що добре відповідає завданням радіолокаційного виявлення, є сигнал цифрового телебачення, який має порівняно високу енергетику та достатньо широкий спектр [1]. У сучасних бістатичних РЛС з телевізійним сигналом освітлення використовується кореляційна обробка прийнятих сигналів. На вхід кореляційного виявлювана поступають комплексні відліки сигналів цільового та опорного каналів (шуми в опорному каналі можливо проігнорувати) [2]:

$$\dot{m}(n) = (\dot{E}_c + \dot{E}_m U_m(\Delta tn - t_d))\exp(j\,\Omega_d \Delta tn) + \dot{h}(n),$$

$$\dot{m}_0(n) = \dot{E}_{c0} + \dot{E}_{m0} U_m(\Delta tn),$$

де n – номер відліку дискретного сигналу; $\dot{E}_c$, $\dot{E}_{co}$ – комплексні амплітуди постійної складової сигналу цифрового телебачення у цільовому та опорному каналах відповідно; $\dot{E}_m$, $\dot{E}_{mo}$ – комплексні амплітуди модуляційної складової сигналу цифрового телебачення у цільовому та опорному каналах відповідно; $U_m(t)$ – нормована за потужністю огинаюча модуляційної складової сигналу цифрового телебачення; Δt – період дискретизації Котельникова, який розраховується за формулою $\Delta t = 1/2f_{max}$, де f_{max} – максимальна частота спектру телевізійного

сигналу; t_d – запізнення відбитого цілю сигналу відносно прямого сигналу; Ω_d – доплеровський здвиг частоти відбитого сигналу; $\dot{h}(n)$ – адитивний шум, який враховує гаусовський шум приймача та завади, викликані розсіянням телевізійного сигналу морською поверхнею при бістатичній радіолокації.

Обмежимся випадком стаціонарної за часом модуляційної складової сигналу цифрового телебачення. Тоді умови нормування та центрування при N відліках дискретного сигналу:

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}U_m^2(\Delta tn+\tau)=1,\ \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}U_m(\Delta tn+\tau)=0, \qquad (1)$$

Дискретна автокореляційна функція модуляційної складової сигналу цифрового телебачення:

$$B_m(\Delta ti)=\frac{1}{N}\sum_{n=1}^{N}U_m(\Delta tn)U_m(\Delta tn-\Delta ti). \qquad (2)$$

Після корекції частоти Доплера на величину Ω_c та віднімання постійної складової прийнятого сигналу у цільовому та опорному каналах при $\Omega_d=\Omega_c$ отримаємо:

$$\dot{m}_m(n)=\dot{E}_m U_m(\Delta tn-t_d)+\dot{h}(n)\exp(-j\,\Omega_c\Delta tn),$$
$$\dot{m}_{m0}(n)=\dot{E}_{m0}U_m(\Delta tn).$$

Теорією розсіяння радіохвиль на тілах складної форми [3] доведено, що амплітуда відбитого цілю фазомодульованого сигналу має нормальне розподілення. Таким чином, амплітуду $\dot{E}_m$ можна вважати комплексною випадковою гаусовською величиною з нульовим середнім значенням та дисперсією $\mu_2\{\dot{E}_m\}=P_m$, де P_m – потужність телевізійного сигналу, відбитого цілю. Потужність сигналу цілі P_m у відповідності до [3 – 5] розраховується за наступною формулою:

$$P_m=\frac{P_{пер}G_{пер}G_{пр}\sigma_ц\lambda^2}{(4\pi)^3R_1^2R_2^2}, \qquad (3)$$

де $P_{пер}$ – потужність випромінювання передатчика, $G_{пер}$ – коефіцієнт підсилення антени передатчика, $G_{пр}$ – коефіцієнт підсилення антени приймача, $\sigma_ц$ – ефективна бістатична площина розсіяння цілі, λ – довжина телевізійної хвилі, R_1 – відстань від передатчика до цілі, R_2 – відстань від приймача до цілі.

Адитивний шум $\dot{h}(n)$ враховує гаусовський шум приймача та завади, викликані розсіянням телевізійного сигналу морською поверхнею при бістатичній радіолокації. Шум приймача являє собою тепловий шум з рівномірним розподілом спектральної щільності потужності N_0, тобто «білий» шум. Завади, викликані розсіянням телевізійного сигналу морською поверхнею при бістатичній радіолокації, описуються теорією розсіяння радіохвиль статистично нерівною поверхнею [6]. Відповідно до цієї теорії електромагнітне поле, розсіяне морською поверхнею, має нормальне «не біле» розподілення. Тоді шум $\dot{h}(n)$ можна розглядати як комплексний випадковий нормальний процес з нульовим середнім значенням та дисперсією $\mu_2\{\dot{h}(n)\}=P_h$, де P_h – потужність шуму. Відношення сигнал-шум за потужністю для виявлення телевізійного сигналу, відбитого цілю, має вигляд

$$\rho_m=P_m/P_h \qquad (4)$$

Згідно з [7] у випадку безперервного фазомодульованого сигналу, яким є телевізійний сигнал, дисперсія P_h дискретизованої напруги адитивного шуму h(n) визначається формулою:

$$P_h = \sigma^2_{мп} + N_0 f_{max} = P_{мп} + N_0 f_{max}, \quad (5)$$

де $\sigma^2_{мп}$ – дисперсія напруги завади, викликаної розсіянням телевізійного сигналу морською поверхнею при бістатичній радіолокації, $\sigma^2_{мп} = P_{мп}$ – потужність цієї завади; N_0 – спектральна щільність теплового гаусовського шуму приймача; f_{max} – максимальна частота спектру телевізійного сигналу.

$P_{мп}$ у відповідності до [3 – 5, 8] розраховується за формулою

$$P_{мп} = \frac{P_{пер} G_{пер} G_{пр} \sigma^0 S \lambda^2}{(4\pi)^3 R_1^2 R_2^2}, \quad (6)$$

де $P_{пер}$ – потужність випромінювання передатчика, $G_{пер}$ – коефіцієнт підсилення антени передатчика, $G_{пр}$ – коефіцієнт підсилення антени приймача, σ^0 – питома ефективна бістатична площина розсіяння морської поверхні, S – площина морської поверхні, що опромінюється, λ – довжина телевізійної хвилі, R_1 – відстань від передатчика до морської поверхні, що опромінюється, R_2 – відстань від приймача до морської поверхні, що опромінюється (рис. 1).

Розрахунок σ^0 можна здійснити на основі теорії розсіяння радіохвиль статистично нерівною поверхнею [6]. Згідно з основним положенням цієї теорії, розсіяння радіохвиль випадковими поверхнями визначається дрібномасштабними складовими хвилювання з хвильовим числом

$$\chi_0 = K(\beta_+ - \alpha_+), \quad (7)$$

де $K = \frac{2\pi}{\lambda}$, α_+ та β_+ – проекції одиничних векторів α та β на горизонтальну площину, θ – напрям хвильового числа χ_0 (рис. 1).

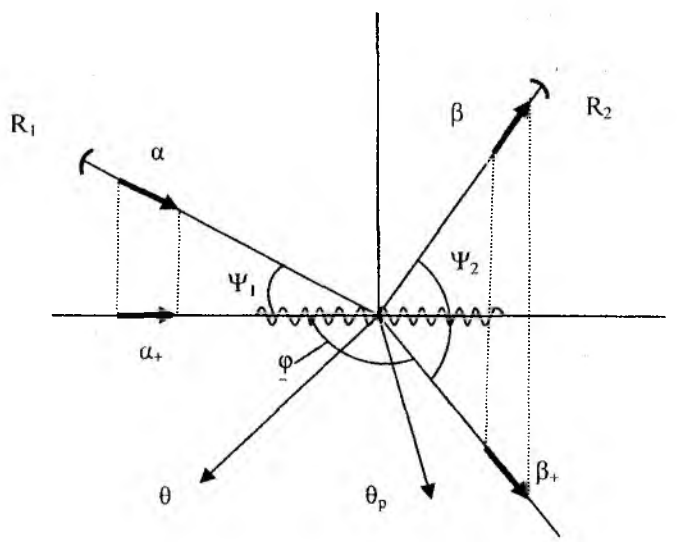


Рис. 1. Геометричні співвідношення при бістатичному розсіянні радіохвиль

Згідно з [8, 9] питома бістатична ефективна площина розсіяння морської поверхні σ^0 є складною функцією декількох елементів:

$$\sigma^0 = f(K, \Psi_1, \Psi_2, \varphi, S(\chi_0)), \tag{8}$$

де Ψ_1, Ψ_2, φ – кути, які відображені на рис. 1; $S(\chi_0)$ – двомірний просторовий спектр складових морського хвилювання з хвильовим числом χ_0, який на ділянці моря з достатньо великою глибиною у випадку повністю розвинутого хвилювання визначається силою, напрямком вітру та тривалістю його дії. $S(\chi_0)$ розраховується у відповідності до [10 – 14].

Результат кореляційної обробки $\dot{W}_i = \sum_{n=1}^{N} \dot{m}_m(n)\dot{m}_{m0}(n-i)$ телевізійного сигналу $\dot{m}_m(n)$, відбитого ціллю, та опорного сигналу $\dot{m}_{mo}(n-i)$ має наступний вигляд [2]:

$$\dot{W}_i = \dot{E}_m \dot{E}_{mo} N B_m(t_d) + \dot{E}_{m0} \sum_{n=1}^{N} \dot{h}(n) U_m(\Delta tn - \Delta ti) \exp(-j\,\Omega_c \Delta tn), \tag{9}$$

де $B_m(t_d)$ – дискретна автокореляційна функція модуляційної складової сигналу цифрового телебачення при $\Delta ti = t_d$ (2)

Рішення про наявність або відсутність сигналу у випадку сигналу з повністю відомими параметрами приймається по величині $\dot{W}_i$ [15]. Величина $\dot{W}_i$ є межею лінійної комбінації нормально розподілених випадкових величин $\dot{E}_m$ та $\dot{h}(n)$, тобто $\dot{W}_i$ є комплексною випадковою гаусовською величиною з нульовим середнім значенням. Її дисперсія $\sigma^2_{W_i} = \overline{\dot{W}_i^2}$. Тоді з урахуванням ρ_m (4) та умови нормування (1) при достатньо великому N [2]:

$$\sigma^2_{W_i} = \overline{\dot{W}_i^2} = P_h \left|\dot{E}_{m_0}\right|^2 N(\rho_m N B_m^2(t_d) + 1). \tag{10}$$

В умовах відсутності сигналу, коли математичне очікування шуму дорівнює 0, математичне очікування випадкової величини $\dot{W}_i$ також дорівнює 0. Тоді умовна щільність ймовірності $P_h(r)$ [15]:

$$P_h(r) = \frac{1}{\sqrt{2\pi}\sigma_h} \exp\left(-\frac{r^2}{2\sigma_h^2}\right). \tag{11}$$

Невипадкова величина σ_h^2, яка входить до (11), є дисперсією випадкової величини $\dot{W}_i = \dot{W}_{hi}$ при наявності лише шуму, тобто при $\rho_m = 0$. Тоді

$$\sigma_h^2 = \overline{\dot{W}_{hi}^2} = P_h \left|\dot{E}_{m_0}\right|^2 N. \tag{12}$$

Виявлення когерентних сигналів з випадковою амплітудою та начальною фазою здійснюється шляхом порівняння з порогом модуля $Z_i = \sqrt{X_i^2 + Y_i^2}$ випадкової величини $\dot{W}_i$ [15]. В умовах наявності лише шуму кожна з незалежних величин X_i та Y_i має нормальне розподілення ймовірності (11) з дисперсіями $\sigma^2_{x_i} = \sigma^2_{y_i} = \frac{1}{2}\sigma_h^2$, тому для Z_i має місце закон розподілення Релєя [15]. Згідно з [16] квадрат релеєвської випадкової величини Z_i^2 має експоненціальне розподілення зі щільністю ймовірності:

$$P_{Z_i^2}(r)=\frac{1}{2\sigma_{x_i}^2}\exp\left(-\frac{r}{2\sigma_{x_i}^2}\right)=\frac{1}{P_h\left|\dot{E}_{m_0}\right|^2N}\exp\left(-\frac{r}{P_h\left|\dot{E}_{m_0}\right|^2N}\right). \quad (13)$$

Ймовірність хибної тривоги в і-му елементі розділення по затримці:

$$F_i=1-\int_0^1 P_{Z_i^2}(r)dr=\exp\left(-\frac{1}{\sigma_h^2}\right)=\exp\left(-\frac{1}{P_h\left|\dot{E}_{m_0}\right|^2N}\right), i\neq 0. \quad (14)$$

Звідси поріг виявлення телевізійного сигналу, відбитого морською ціллю при бістатичній радіолокації $l_{opt,i}$, оптимальний за критерієм Неймана – Пірсона для завданої ймовірності F_r:

$$\begin{aligned} l_{opt,i}&=-\sigma_h^2\ln F_r=-\overline{\dot{W}_{hi}^2}\ln F_r=-P_h\left|\dot{E}_{m_0}\right|^2N\ln F_r= \\ &=-\left(\frac{P_{пер}G_{пер}G_{пр}\sigma^0S\lambda^2}{(4\pi)^3R_1^2R_2^2}+N_0f_{max}\right)\left|\dot{E}_{m_0}\right|^2N\ln F_r, i\neq 0. \end{aligned} \quad (15)$$

Із співвідношень (15) видно, що на практиці для визначення оптимального порогу виявлення цілі достатньо здійснити розрахунки за формулою:

$$l_{opt,i}=-\overline{\dot{W}_{hi}^2}\ln F_r, i\neq 0, \quad (16)$$

де $\overline{\dot{W}_{hi}^2}$ – усереднене значення квадратів результатів кореляційної обробки телевізійного сигналу в і-му елементі розділення по дальності, отриманих з попередніх замірів в умовах відсутності сигналу цілі.

При наявності в елементі розділення по затримці ($|i-t_d/\Delta t|\le 0$), який аналізується, сигналу з випадковою амплітудою та начальною фазою, для Z_i^2 також має місце експоненціальний закон розподілення [16]:

$$\begin{aligned} P_{Z_i^2}(r)&=\frac{1}{2\sigma_{x_i}^2}\exp\left(-\frac{r}{2\sigma_{x_i}^2}\right)=\frac{1}{\sigma_{\dot{W}_i}^2}\exp\left(-\frac{r}{\sigma_{\dot{W}_i}^2}\right)= \\ &=\frac{1}{P_h\left|\dot{E}_{m_0}\right|^2N(\rho_mNB_m^2(t_d)+1)}\exp\left(-\frac{r}{P_h\left|\dot{E}_{m_0}\right|^2N(\rho_mNB_m^2(t_d)+1)}\right). \end{aligned} \quad (17)$$

Ймовірність правильного виявлення при оптимальному порозі:

$$D_i=1-\int_0^{l_{opt,i}}P_{Z_i^2}(r)dr=\exp\left(-\frac{l_{opt,i}}{\sigma_{\dot{W}_i}^2}\right)=F_r^{\sigma_h^2/\sigma_{\dot{W}_i}^2}=F_r^{\overline{\dot{W}_{hi}^2}/\overline{\dot{W}_i^2}}=F_r^{1/(\rho_mNB_m^2(t_d)+1)}=F_r^{1/\left(\frac{P_m}{P_h}NB_m^2(t_d)+1\right)}, i\neq 0. \quad (18)$$

випадку високої кореляції відбитого цілю сигналу з опорним, коли $B_m(t_d)\approx 1$, формулу для розрахунку D_i можливо записати у наступному вигляді:

$$D_i=F_r^{\sigma_h^2/\sigma_{\dot{W}_i}^2}=F_r^{\overline{\dot{W}_{hi}^2}/\overline{\dot{W}_i^2}}=F_r^{1/\left(\frac{P_m}{P_h}N+1\right)}, i\neq 0. \quad (19)$$

Наведені вище співвідношення дозволяють здійснювати розрахунки характеристик виявлення сигналів, відбитих об'єктами, розташованими на морській поверхні, бістатичною радіолокаційною станцією з цифровим телевізійним сигналом освітлення. Зазначені розрахунки базуються на моделях статистичної радіотехніки. Їх специфікою є те, що вони врахо-

ують особливості кореляційної обробки цифрового телевізійного сигналу і особливості роз-іяння радіохвиль морською поверхнею.

Результати, отримані у ході досліджень, можуть використовуватись для побудови алго-итмів прийняття оптимального рішення у системах виявлення об'єктів на морі.

Список літератури: 1. *Семашко, П.Г., Пархоменко, Н.Г., Охрименко, А.Е.* Перспективы полуактив-ной радиолокации в связи с развитием служб цифрового радиовещания // Успехи современной радио-электроники. – М., 2010. – №7. – С. 38–46. 2. *Семашко, П.Г., Охрименко, А.Е.* Характеристики обна-ружения телевизионного сигнала яркости в радиолокационной системе с посторонним подсветом // Доклады БГУИР. – Минск : Белорус. гос. ун-тет информатики и радиоэлектроники, 2005. – №2. – С.60–64. 3. *Штагер, Е.А.* Рассеяние радиоволн на телах сложной формы. – М. : Радио и связь, 1986. – 184 с. 4. *Сколник, М.* Справочник по радиолокации. – Т. 1. Основы радиолокации. – М. : Сов. радио, 1976. – 456 с. 5. *Черняк, В.С.* Многопозиционная радиолокация. – М. : Радио и связь, 1993. – 416 с. 6. *Басс, Ф.Г., Фукс, М.И.* Рассеяние волн на статистически неровной поверхности. -М. : Наука, 1972. – 424 с. 7. *Ширман, Я.Д., Манжос, В.Н.* Теория и техника обработки радиолокационной информации на фоне помех. – М. : Радио и связь, 1981. – 416 с. 8. *Бондур, В.Г., Доброзраков, А.Д., Курекин, А.С., Ку-рекин, А.А., Пичугин, А.П., Яцевич, С.Е.* Рассеяние радиоволн морской поверхностью при бистатичес-кой локациинич // Исследование Земли из космоса. – М., 2009. – №6. – С. 3–15. 9. *Калмыков, А.И., Ку-рекин, А.С., Островский, И.Е., Пустовойтенко, В.В.* Двухпозиционное рассеяние радиоволн поверх-ностью моря при малых углах скольжения // Известия вузов. Радиофизика. – 1966. – Т.10. – № 6. – С. 1073–1082. 10. *Рытов, С.М.* Введение в статистическую радиофизику. – М. : Наука, 1966. – 404 с. 11. *Филипс, О.М.* Динамика верхнего слоя океана : пер. с англ. – М. : Мир, 1969. – 420 с. 12. *Полни-ков, В.Г.* Нелинейная теория случайного поля волн на воде. – М. : ЛЕНАНД, 2007. – 408 с. 13. *Лобко-ва, Л.М.* Распространение радиоволн над морской поверхностью. М. : Радио и связь, 1991. 255 с. 14. *Pierson W.J., Moskowits L.* A proposed spectrum for fully developed wind seas based on the similarity theory of S.A. Kitaigorodskii // J. Geophys. Res. – 1964. – 69. № 24. – P. 5181–5190. 15. *Ширман, Я.Д.* Теоретические основы радиолокации. – М. : Сов. радио, 1970. – 560 с. 16. *Левин, Б.Р.* Теоретические основи статистической радиотехники. – М. : Сов. радио, 1974. – 552 с. 17. *Saini, R., Cherniakov, M.* Investigation of digital TV-terrestrial signal for radar application // Proc. of the XI European signal processing conference (EUSIPCO), Toulouse, France, 3 – 6 Sept. 2002. Toulouse,2002. V. 3. P. 15–18. 18. *Griffiths, H., Long, N.* Television-Based Bistatic Radar // IEE Proceedings. – 1986. – V. 13. – P.649–657. 19. *Poullin, D., Lesturgie, M.* Radar Multistatic a Emissions Non Cooperatives // IEE Conference Publications (Radar). – 1994. – P.370–375. 20. *Howland, P.* Target Tracking Using Television-Based Bistatic Radar // IEE Proceed-ings on Radar, Sonar and Navigation. – 1999. – №146(3). – P.166–174. 21. *Ching-Wei Wesley Chang.* Sys-tem Level Investigations of Television Based Bistatic Radar // Dissertation – University of Cape Nown. – 2005. – 107 p.

Академія військово-морських сил
імені П.С. Нахімова

Надійшла до редколегії 14.09.2012

УДК 621.396.2

А.А. БУЛАНЫЙ, Г.В. МАЙСТРЕНКО, А.А. СТРЕЛЬНИЦКИЙ, канд. техн. наук,
В.М. ШОКАЛО, д-р техн. наук

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ПОМЕХОЗАЩИЩЕННОСТИ И СПЕКТРАЛЬНОЙ ЭФФЕКТИВНОСТИ WI-FI КАНАЛОВ СВЯЗИ С ЛИНЕЙНЫМИ И ДВУМЕРНЫМИ АДАПТИВНЫМИ АНТЕННЫМИ РЕШЕТКАМИ ПРИ ВОЗДЕЙСТВИИ НЕСКОЛЬКИХ ПОМЕХ И ОДНОГО СИГНАЛА

Введение

Исследованию каналов связи с адаптивными антенными решетками (ААР) посвящено много работ (например, [1 – 4]), в которых описаны алгоритмы адаптации, а также схемы адаптивных антенн. Однако в этих работах пока не проведены детальные исследования малоэлементных ААР, которые нашли применение в рэлеевских Wi-Fi каналах связи. В частности, не изучены уровни помехозащищенности и производительности Wi-Fi каналов связи с малоэлементными линейными и двумерными адаптивными антенными решетками при воздействии нескольких помех и флуктуации направления их прихода. Частичная ликвидация этого пробела для частного случая прихода сигнала с направления близко к нормали ААР и является целью данной статьи.

Постановка и решение задачи

Проведем исследование помехозащищенности и спектральной эффективности Wi-Fi каналов связи с линейными и двумерными ААР при воздействии нескольких помех и одного сигнала. Для этого будем использовать алгоритм, разработанный авторами для рэлеевских Wi-Fi каналов связи в [5]. Все расчеты проведем для случая изотропных излучателей и при условии, что мощность тепловых шумов по отношению к мощности помехи очень мала и ею можно пренебречь. При этом будем считать, что уровень помехозащищенности оценивается вероятностью битовой ошибки BER. Значение BER в рэлеевском канале связи для бинарной модуляции вычисляется по формуле [6]:

$$BER = \frac{1}{2}\left(1 - \sqrt{\frac{\rho}{\rho+1}}\right),$$

где ρ – отношение сигнал-помеха (ОСП) рэлеевского канала связи на входе приемника.

Спектральная эффективность $C_э$ рэлеевского канала определяется выражением [6]

$$C_э = -\frac{1}{\ln 2} e^{\frac{1}{\rho}} \int_{-\infty}^{\frac{1}{\rho}} \frac{e^t}{t} dt.$$

В качестве ОСП ρ будем принимать отношение мощностей приходящих полезного и помехового сигналов, умноженное на отношение значений полученной оптимальной диаграммы направленности (ДН) по мощности в направлении полезного и помехового сигналов

$$\rho = \frac{P_C}{P_П} \cdot \frac{F_C}{\sum_j F_{Пj}},$$

где $\frac{P_C}{P_П}$ – отношение мощностей сигнала (P_C) и помехи ($P_П$); $F_C = F(\Theta_C, \varphi_C)$ – значение ДН по мощности в направлении полезного сигнала; $\sum_j F_{Пj}$ – сумма по всем помехам (j=1,2,3 значений ДН, усредненных по секторам флуктуации помеховых сигналов.

Так как в разработанном алгоритме направление прихода помехи является случайной величиной с известной плотностью распределения, то значение ДН в направлении помехового сигнала также является случайной величиной и его среднее значение вычисляется по формуле

$$F_{П} = \int_{\Theta_{П}-\delta}^{\Theta_{П}+\delta} \int_{\varphi_{П}-\delta}^{\varphi_{П}+\delta} F(\varphi,\theta)\cdot \mathrm{g}(\varphi,\theta)\, d\varphi\, d\theta ,$$

где $\mathrm{g}(\varphi,\theta)$ – плотность распределения направления прихода помехи в секторе 2δ.

Численные исследования

Исследования проводились на четырехэлементных антенных решетках – линейной и квадратной. Изменения формы ДН в зависимости от количества помех приведены на рис. 1.

Расчеты выполнялись для решеток с шагом d/λ=0,7. ДН линейной антенны изображены на рис.1, *а – в*, квадратной – рис.1, *г – е*.

Во всех примерах направление прихода сигнала $\Theta_С$ было принято 80°, а направления прихода помех $\Theta_П$ были выбраны: 20° – для случая одной помехи; 20° и 135° – для случая двух помех; 20°, 135° и 40° – для случая трех помех. Ширина секторов флуктуации помех одинакова для всех помех и равна ±3°.

ДН адаптивной антенны, рассчитанные в присутствии помех, изображены на рис.1 сплошной линией, а штрих-пунктиром показаны оптимальные по максимуму ОСП ДН ААР без учета влияния помех.

Как видно из рисунков, форма ДН для данного угла прихода сигнала не существенно зависит от количества помех, и для того чтобы понять где хуже «работает» алгоритм адаптации, составлены табл. 1 – 3, в которых представлены итоговые параметры ААР. Проанализируем их.

Максимумы нормированных ДН по мощности в направлении прихода сигнала $F(\Theta_С)$ в линейной и квадратной решетках отличаются менее чем на два порядка. Зато минимумы ДН в направлении прихода помех в случае одной помехи (см.табл.1) для квадратной ААР на четыре и более порядков меньше, чем для линейной ААР. Этим и определяется выигрыш в отношении сигнал/помеха ρ для квадратной антенны по сравнению с линейной. В свою очередь увеличение ρ приводит к существенному улучшению параметров BER и СЭ в квадратной решетке. Из табл. 1 следует, что квадратная ААР обеспечивает высокое качество передачи информации (BER < -40 *дБ*) при любом шаге решетки d/λ. С помощью же линейной ААР реализуется только среднее качество с BER < -20 *дБ*, при этом производительность канала уменьшается практически в два с половиной раза по сравнению со случаем применения квадратной ААР.

Таблица 1

Тип антенн	Линейная ААР	Квадратная ААР	Линейная ААР	Квадратная ААР	Линейная ААР	Квадратная ААР	Линейная ААР	Квадратная ААР
d/λ	0.3		0.5		0.7		1.0	
$F_{С}$	0.95	0.02	1	0.41	1	0.20	0.73	0.11
$F_{П1}$	9.32×10^{-4}	1.11×10^{-9}	8.12×10^{-4}	2.62×10^{-8}	8.81×10^{-4}	1.90×10^{-8}	9.70×10^{-4}	5.46×10^{-8}
ρ, дБ	30.07	72.52	30.90	71.99	30.54	70.27	28.75	62.86
BER, дБ	-36.09	-78.54	-36.93	-78.01	-36.56	-76.29	-34.78	-68.88
$C_{Э}$, $\frac{\text{бит}}{\text{с}\cdot\text{Гц}}$	9.17	23.26	9.44	23.08	9.32	22.51	8.73	20.05

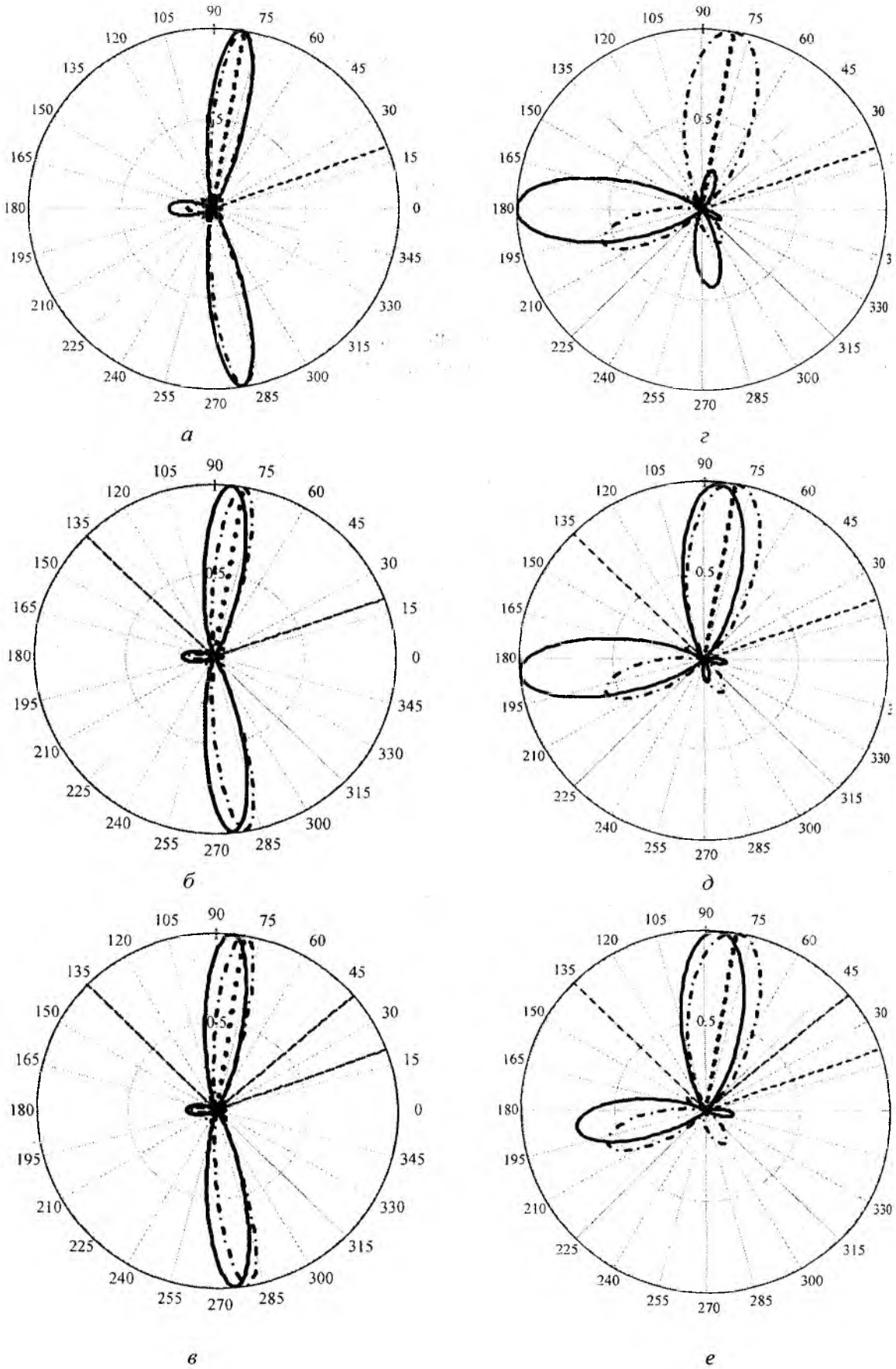


Рис.1. ДН линейной и квадратной ААР с шагом решетки d/λ=0.7 при направлении полезного сигнала Θ_c = 80° и разного количества помех (одной помехи – *а*, *г*, двух помех – *б*, *д*, трех помех – *в*, *е*)

Ситуация существенно изменяется при воздействии двух (табл. 2) и трех (табл. 3) помех. Как видно из табл. 2, 3, с ростом числа помех значения $F_П$ увеличиваются, что приводит к уменьшению величины BER, особенно в квадратной ААР. При этом линейная и квадратная ААР обеспечивают достижение только среднего качества передачи информации.

Таблица 2

Тип антенн	Линейная ААР	Квадратная ААР	Линейная ААР	Квадратная ААР	Линейная ААР	Квадратная ААР	Линейная ААР	Квадратная ААР
d/λ	0.3		0.5		0.7		1.0	
F_C	0.98	0.25	1	0.84	0.93	0.94	0.10	0.61
$F_{П1}$	1.62×10^{-3}	1.14×10^{-3}	9.30×10^{-4}	5.99×10^{-4}	9.27×10^{-4}	1.08×10^{-3}	4.91×10^{-3}	9.59×10^{-3}
$F_{П2}$	1.16×10^{-3}	1.76×10^{-4}	8.97×10^{-4}	6.57×10^{-4}	9.33×10^{-4}	1.25×10^{-5}	2.77×10^{-3}	8.22×10^{-4}
ρ, дБ	29.29	22.81	28.38	28.28	27.00	29.35	11.05	17.67
BER, дБ	-31.51	-28.85	-33.41	-34.30	-33.03	-35.37	-17.32	-23.75
$C_Э$, $\frac{\text{бит}}{\text{с}\cdot\text{Гц}}$	7.66	6.79	8.28	8.58	8.16	8.93	3.19	5.15

Таблица 3

Тип антенн	Линейная ААР	Квадратная ААР	Линейная ААР	Квадратная ААР	Линейная ААР	Квадратная ААР	Линейная ААР	Квадратная ААР
d/λ	0.3		0.5		0.7		1.0	
F_C	0.67	0.04	0.80	0.81	0.92	0.94	0.39	0.24
$F_{П1}$	1.93×10^{-3}	1.57×10^{-5}	9.00×10^{-4}	2.34×10^{-4}	9.31×10^{-4}	1.23×10^{-3}	5.38×10^{-3}	9.88×10^{-4}
$F_{П2}$	6.01×10^{-3}	5.71×10^{-4}	1.17×10^{-3}	1.81×10^{-3}	9.27×10^{-4}	1.03×10^{-3}	0.03	5.75×10^{-3}
$F_{П3}$	1.92×10^{-3}	2.80×10^{-5}	9.62×10^{-4}	4.69×10^{-4}	9.15×10^{-4}	2.07×10^{-3}	7.17×10^{-3}	1.23×10^{-3}
ρ, дБ	18.29	17.96	24.20	25.07	25.22	23.36	9.65	14.72
BER, дБ	-24.36	-24.03	-30.23	-31.10	-31.25	-29.40	-16.00	-20.85
$C_Э$, $\frac{\text{бит}}{\text{с}\cdot\text{Гц}}$	5.34	5.24	7.24	7.52	7.57	6.97	2.81	4.25

Так как приведенные в табл.2,3 данные получены при ОСП приходящих сигналов равном единице, данные для BER при других значениях ОСП на входе ААР сведены в табл. 4.

С ростом ОСП значения BER улучшаются и можно определить те значения ОСП на входе решетки, при которых обеспечивается не только среднее и высокое качество передачи информации в канале.

Важно знать зависимость BER от ширины зоны флуктуации помехи 2δ. Эти графики изображены на рис. 2 (при воздействии одной помехи), на рис. 3 (при воздействии двух помех) и на рис. 4 (при воздействии трех помех). Квадратная решетка явно выигрывает в случае одной помехи. При этом высокое качество передачи информации обеспечивается до величины $\delta\approx17^\circ$ (см. рис. 2). Линейная же ААР позволяет реализовать только среднее качество передачи информации (BER > -40*дБ*) во всем диапазоне углов δ от 1° до 20°.

Таблица 4

Линейная ААР					Квадратная ААР			
ОСП / d/λ	1	10	100	1000	1	10	100	1000
При воздействии одной помехи								
0,3	-36.09	-46.09	-56.09	-66.09	-78.54	-88.54	-98.54	-108.54
0,5	-36.93	-46.93	-56.93	-66.93	-78.01	-88.01	-98.01	-108.01
0,7	-36.56	-46.56	-56.56	-66.56	-76.29	-86.29	-96.29	-106.29
1,0	-34.78	-44.77	-54.77	-64.77	-68.88	-78.88	-88.88	-98.88
При воздействии двух помех								
0,3	-31.51	-41.50	-51.50	-61.50	-28.85	-38.84	-48.84	-58.84
0,5	-33.41	-43.41	-53.41	-63.41	-34.30	-44.30	-54.30	-64.30
0,7	-33.03	-43.02	-53.02	-63.02	-35.37	-45.37	-55.37	-65.37
1,0	-17.32	-27.10	-37.01	-47.01	-23.75	-33.70	-43.69	-53.69
При воздействии трёх помех								
0,3	-24.36	-34.32	-44.31	-54.31	-24.03	-33.99	-48.98	-53.98
0,5	-30.23	-40.22	-50.22	-60.22	-31.10	-41.09	-51.09	-61.09
0,7	-31.25	-41.24	-51.24	-61.24	-29.40	-39.38	-49.38	-59.38
1,0	-16.00	-25.7	-35.67	-45.68	-20.85	-30.76	-40.75	-50.74

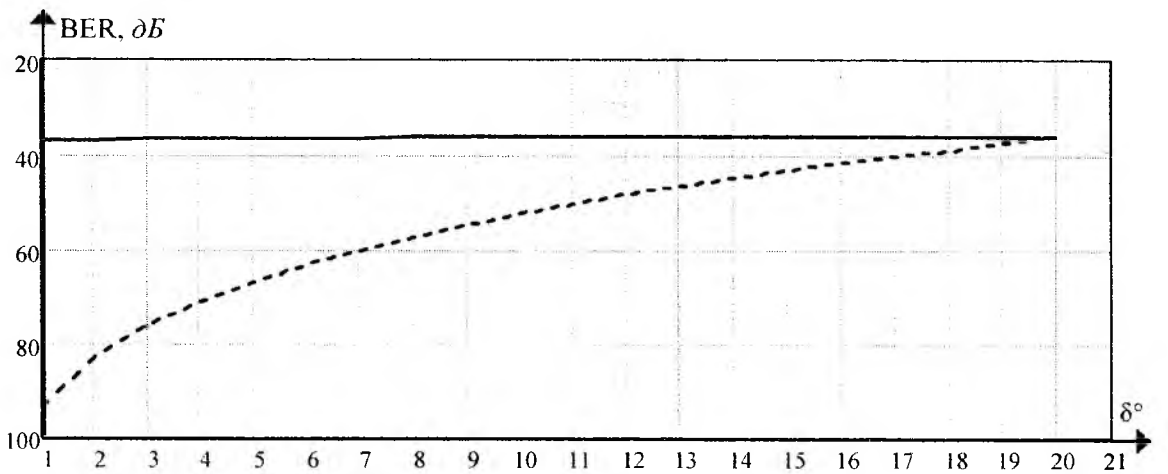


Рис.2. Зависимость BER для линейной ААР (сплошная линия) и квадратной ААР (пунктирная линия при воздействии одной помехи

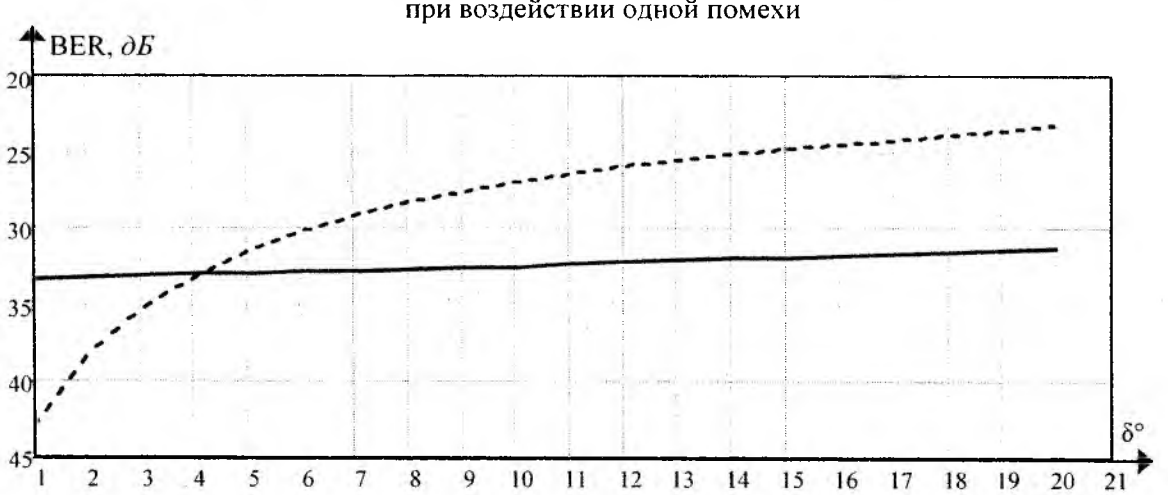


Рис. 3. Зависимость BER для линейной ААР (сплошная линия) и квадратной ААР (пунктирная линия при воздействии двух помех

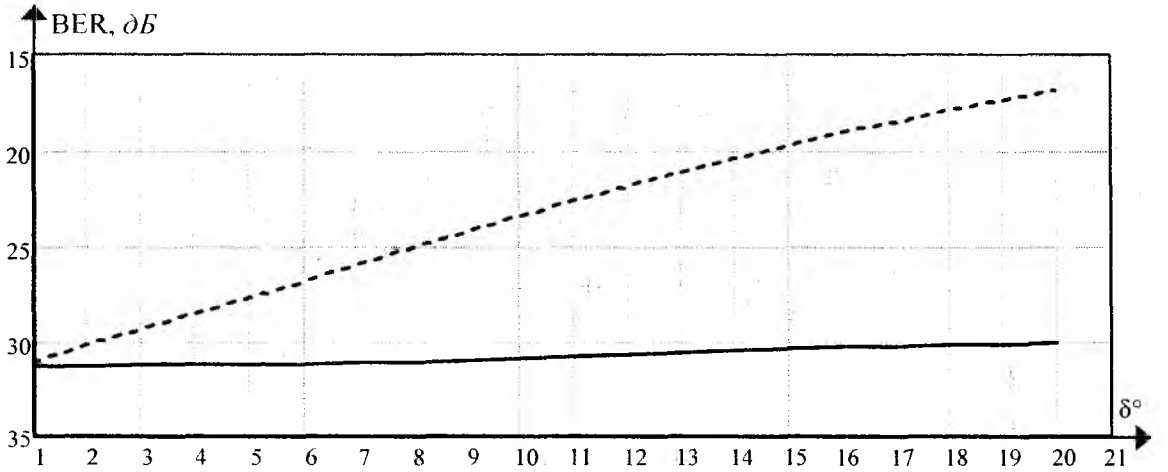


Рис.4. Зависимость BER для линейной ААР (сплошная линия) и квадратной ААР (пунктирная линия) при воздействии трех помех

Как уже отмечалось, при воздействии двух помех канал связи с квадратной ААР имеет величину BER соответствующую среднему качеству передачи информации (см.рис.3) и при $\delta > 4^\circ$ проигрывает линейной ААР, которая позволяет достичь при $1^\circ \le \delta \le 20^\circ$ качество передачи информации не хуже – 33 *дБ*. В случае же трех помех, величина BER еще ухудшается и для квадратной ААР при значении сектора флуктуации помехи δ от 15° до 20° соответствует только низкому качеству передачи информации. Ситуация может быть улучшена только за счет увеличения ОСП более чем в 10 раз.

Сопоставляя достоинства и недостатки линейных и двумерных ААР, необходимо обсудить вопрос о сложности их технической реализации. Для этого следует сравнить амплитудно-фазовые распределения (АФР) обеих решеток (см.табл.5,6). В таблицах через A_i обозначены значения нормированных амплитуд на входах излучателей а через φ_i – значения фаз. Нетрудно сделать вывод, что оба варианта ААР одинаковы по сложности реализации, т.к. и в том, и в другом случае требуется обеспечение не только заданного фазового, но и амплитудного распределения.

Таблица 5

d/λ	A_1	A_2	A_3	A_4	φ_1	φ_2	φ_3	φ_4
При воздействии одной помехи								
0,3	0.95	1	1	0.95	4,32	0	22,70	18,38
0,5	0.64	1	1	0.64	0	15,36	70,70	86.06
0,7	0.48	1	1	0.48	0	44,32	76,06	123,38
1,0	0.86	1	1	0.86	232,57	0	40,26	163,69
При воздействии двух помех								
0,3	1	0.98	0.98	1	337,00	54,35	0	77,35
0,5	0.39	1	1	0.39	0	32,22	61,32	93,55
0,7	0.36	1	1	0.36	0	26,33	53,24	76,56
1,0	0.85	1	1	0.85	0	277,91	129,48	47,39
При воздействии трёх помех								
0,3	1	0.98	0.98	1	0	267,89	27,47	295,35
0,5	0.43	1	1	0.43	0,02	0	3,75	3,73
0,7	0.37	1	1	0.37	0	24,76	50,58	75,36
1,0	1	0.81	0.81	1	0	142,52	31,92	174,43

Таблица 6

d/λ	A_1	A_2	A_3	A_4	φ_1	φ_2	φ_3	φ_4
При воздействии одной помехи								
0,3	1	0.85	1	0.85	147.88	60.97	0	86.91
0,5	1	0.72	1	0.72	36.39	0	10.19	46.58
0,7	1	0.64	1	0.64	282.33	288.91	6.59	0
1,0	1	0.57	1	0.57	183.27	245.31	62.04	0
При воздействии двух помех								
0,3	1	0.89	1	0.89	141.97	0	84.83	226.80
0,5	1	0.90	1	0.90	105.66	0	41.39	147.04
0,7	0.83	1	0.83	1	322.76	286.51	323.75	0
1,0	0.27	1	0.27	1	264.90	0	140.76	45.66
При воздействии трёх помех								
0,3	1	0.97	1	0.97	176.02	72.21	0	103.81
0,5	1	0.96	1	0.96	63.09	6.38	0	56.71
0,7	0.85	1	0.85	1	332.32	278.04	305.72	0
1,0	1	0.85	1	0.85	149.47	268.57	69.05	0

Выводы

Для частного случая прихода сигнала с направления близко к нормали ААР получены новые результаты сравнительного анализа помехозащищенности и спектральной эффективности Wi-Fi каналов связи с малоэлементными линейными и квадратными ААР, которые заключаются в следующем:

- показано, что при воздействии одной помехи на ААР канал связи с квадратной решеткой существенно выигрывает по параметрам BER и СЭ по сравнению со случаем использования линейной ААР за счет реализации более глубоких нулей ДН в направлении прихода помехи;
- предложена модель расчета BER и СЭ при воздействии нескольких помех, основанная на предположении, что помехи являются аддитивными, они имеют одинаковый равномерный закон распределения прихода помехи в секторе $(\Theta_П \pm \delta)$, $1° \le \delta \le 20°$.
- с использованием предложенной модели расчетным путем выяснено, что в случае воздействия нескольких помех нули в ДН квадратных ААР "заплывают" больше, чем в линейных, что приводит к более плохим результатам по качеству передаваемой информации и производительности канала.

Список литературы: 1. *Уидроу, Б., Стирнз, С.* Адаптивная обработка сигналов : пер. с англ. – М. : Радио и связь, 1989. – 440 с. 2. *Титаренко, Л.А.* Адаптивная пространственная обработка сигналов в условиях априорной неопределенности. – Х. : ХНУРЭ; Коллегиум, 2004. – 216с. 3. *Быховский, М.А.* Приём широкополосных сигналов с помощью адаптивных антенн // Мобильные системы. – 2006. – №8. – С. 21-27. 4. *Дурманов, М.А., Широков, И.Б.* Адаптивная антенная система базовой станции стандарта GSM 900 // Электроника и связь. – 2009. – Вып. 1(48). – С. 76-81. 5. *Майстренко, Г.В., Рыбалко, А.М., Стрельницкий, А.А., Шокало, В.М.* Влияние случайного изменения направления прихода помехи на помехозащищенность Wi-Fi канала связи с адаптивными антеннами // Радиотехника. – 2012. – Вып. 169. – С. 168-175. 6. *Ермолаев, В.Т., Флаксман, А.Г.* Теоретические основы обработки сигналов в системах мобильной радиосвязи. – Режим доступа: http://www.unn.ru/pages/e-library/methodmaterial/files/56.pdf .

Харьковский национальный университет радиоэлектроники

Поступила в редколлегию 04.09.2012

УДК 621.376

С.В. МАРЧЕНКО, В.М.МОРОЗОВ , канд. физ.-мат. наук,
А.М. СЪЯНОВ, д-р техн. наук

ИССЛЕДОВАНИЕ ФАР С ДИЭЛЕКТРИЧЕСКИМ ЗАПОЛНЕНИЕМ И СОГЛАСУЮЩЕЙ ПЕРИОДИЧЕСКОЙ СТРУКТУРОЙ

Введение

Существуют различные методы согласования ФАР: применение диэлектрических вставок и слоев, диафрагм [1], индуктивных штырей, импедансных фланцев [2] и т.д. В данной работе рассматривается согласование за счет применения согласующей периодической структуры (СПС) совместно с диэлектрическими слоями и вставками.

Постановка задачи

На рис.1 представлена единичная ячейка бесконечной линейной волноводной ФАР, расположенная в начале координат. Рассматривается случай сканирования в Н-плоскости. Согласующая периодическая структура имеет те же размеры поперечного сечения, что и основная решетка.

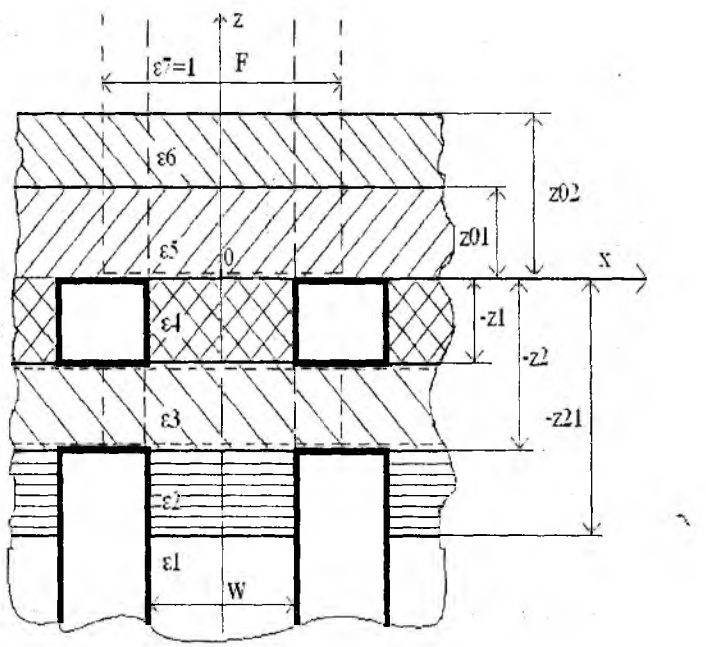


Рис.1. Единичная ячейка ФАР с согласующим устройством и диэлектрическим заполнением

Вся область определения электромагнитного поля разбита на две частичные и одну пронизывающую область:

I (пронизывающая область) – регулярный волновод, продлённый в бесконечность с диэлектрическим заполнением (ε1-ε7):

$$\frac{-W}{2} \le x \le \frac{W}{2};\ -\infty < z < +\infty$$

II (частичная область) –"канал Флоке" конечной длины, заполненный диэлектриком (ε3):

$$\frac{-F}{2} \le x \le \frac{F}{2};\ -z2 \le z \le -z1.$$

III (частичная область) – полубесконечный “канал Флоке”, заполненный диэлектриками (ε5-ε7) :

$$\frac{-F}{2} \le x \le \frac{F}{2}; \quad 0 \le z < +\infty.$$

Решение задачи

На основании подхода, изложенного в [3], получим интегральное представление для определения полного поля пронизывающей области. Учет диэлектрического заполнения в функции Грина при её построении и в стороннем источнике поля описан в работе [4]. В записанном ниже интегральном представлении учтено равенство полей принадлежащих различным областям на соответствующих границах:

$$\begin{aligned}
&E_y^{I(n)}(x,z) = E_{y_{exc}}^{I(n)}(x,z) + \\
&+\int_{-z2}^{-z1}\left\{E_y^{II(3)}(x',z')\frac{\partial G_3^I(x,z;x',z')}{\partial x'}\bigg|_{x'=\frac{W}{2}} - E_y^{II(3)}(x',z')\frac{\partial G_3^I(x,z;x',z')}{\partial x'}\bigg|_{x'=\frac{W}{2}}\right\}dz' + \\
&+\int_{0}^{z01}\left\{E_y^{III(5)}(x',z')\frac{\partial G_5^I(x,z;x',z')}{\partial x'}\bigg|_{x'=\frac{W}{2}} - E_y^{III(5)}(x',z')\frac{\partial G_5^I(x,z;x',z')}{\partial x'}\bigg|_{x'=\frac{W}{2}}\right\}dz' \\
&+\int_{z01}^{z02}\left\{E_y^{III(6)}(x',z')\frac{\partial G_6^I(x,z;x',z')}{\partial x'}\bigg|_{x'=\frac{W}{2}} - E_y^{III(6)}(x',z')\frac{\partial G_6^I(x,z;x',z')}{\partial x'}\bigg|_{x'=\frac{W}{2}}\right\}dz' + \\
&+\int_{z02}^{+\infty}\left\{E_y^{III(7)}(x',z')\frac{\partial G_7^I(x,z;x',z')}{\partial x'}\bigg|_{x'=\frac{W}{2}} - E_y^{III(7)}(x',z')\frac{\partial G_7^I(x,z;x',z')}{\partial x'}\bigg|_{x'=\frac{W}{2}}\right\}dz' \qquad (1) \\
&x \in \left[-\frac{W}{2},\frac{W}{2}\right]; \quad x' = \pm\frac{W}{2}; \quad z' \in [-z1,z2] \cup [0,z01] \cup [z01,z02] \cup [z02,+\infty)
\end{aligned}$$

где $E_y^{I(n)}(x,z)$ – тангенциальная компонента напряженности электрического поля пронизывающей области n-подобласти (n=1÷7); $E_{y_{exc}}^{I(n)}(x,z)$ – тоже для стороннего источника $E_y^{II(3)}(x,z)$ – тангенциальная компонента напряженности электрического поля II области (n=3); $E_y^{III(5)}(x,z), E_y^{III(6)}(x,z), E_y^{III(7)}(x,z)$ – тангенциальная компонента напряженности электрического поля II области n-подобласти (n=5÷7); $G_n^I(x,z;x',z')$ – функция Грина бесконечного волновода с плоскослоистым диэлектрическим заполнением (n=1÷7).

Продольные компоненты определяются с учетом месторасположения точек источников поля и наблюдения в интегральном представлении и свойств симметрии функции Грина:

$$f^{[n]}_{MWG}(z,z') = \begin{cases} kIr\cdot e^{CJ\cdot CWL^{[n]}_{MWG}(z+z21)}, & z\in(-\infty;-z21],\quad n=1,\quad z\in(-\infty;-z21] \\ kIIt\cdot e^{-CJ\cdot CWL^{[n]}_{MWG}(z+z2)} + kIIr\cdot e^{CJ\cdot CWL^{[n]}_{MWG}(z+z2)} + & \\ \dfrac{e^{CJ\cdot CWL^{[n]}_{MWG}(z+z21)}}{CJ\cdot CWL^{[n]}_{MWG}}, & z'\in[-z21;-z2]\quad n=2,\quad z\in[-z21;-z2] \\ kIIIt\cdot e^{-CJ\cdot CWL^{[n]}_{MWG}(z+z1)} + kIIIr\cdot e^{CJ\cdot CWL^{[n]}_{MWG}(z+z1)} + & \\ +\dfrac{e^{CJ\cdot CWL^{[n]}_{MWG}(z+z21)}}{CJ\cdot CWL^{[n]}_{MWG}}, & z'\in[-z2;-z1]\quad n=3,\quad z\in[-z2;-z1] \\ kIVt\cdot e^{-CJ\cdot CWL^{[n]}_{MWG}z} + kIVr\cdot e^{CJ\cdot CWL^{[n]}_{MWG}z} + & \\ +\dfrac{e^{CJ\cdot CWL^{[n]}_{MWG}(z+z21)}}{CJ\cdot CWL^{[n]}_{MWG}}, & z'\in[-z1;0]n=4,\quad z\in[-z1;0] \\ kVt\cdot e^{-CJ\cdot CWL^{[n]}_{MWG}(z+z01)} + kVr\cdot e^{CJ\cdot CWL^{[n]}_{MWG}(z+z01)} & z\in[0;z01]\quad n=5 \\ kVIt\cdot e^{-CJ\cdot CWL^{[n]}_{MWG}(z+z02)} + kVIr\cdot e^{CJ\cdot CWL^{[n]}_{MWG}(z+z02)} & z\in[z01;z02]\quad n=6 \\ kVIIt\cdot e^{-CJ\cdot CWL^{[n]}_{MWG}(z+z02)} & z\in[z02;+\infty)\quad n=7 \end{cases}$$

"Сшивая" выражения для продольных компонент и их производных на границах раздела z=-zI; 0; zII, находим функцию Грина первой области.

Фиксируя точки наблюдения в интегральном представлении при z=-z2; -z1; 0 и учитывая равенство тангенциальных компонент векторов напряженностей областей на этих границах, получим систему интегральных уравнений для полей II и III областей. Напряженности электрического поля II и III областей представлены в виде рядов Фурье:

$$E^{II(3)}_{y}(x,z) = \sum_{nf=-\infty}^{\infty}\left(T3_{nf}e^{-CJ\cdot\gamma3_{nf}\cdot(z+z1)} + R3_{nf}e^{CJ\cdot\gamma3_{nf}(z+z1)}\right)\cdot FD_{nf}(x)\quad,$$

$$E^{III(5)}_{y}(x,z) = \sum_{nf=-\infty}^{\infty}\left(T5_{nf}e^{-CJ\cdot\gamma5_{nf}\cdot(z-z01)} + R5_{nf}e^{CJ\cdot\gamma5_{nf}(z-z01)}\right)\cdot FD_{nf}(x)\quad,$$

$$E^{III(6)}_{y}(x,z) = \sum_{nf=-\infty}^{\infty}\left(T6_{nf}e^{-CJ\cdot\gamma6_{nf}\cdot(z-z02)} + R6_{nf}e^{CJ\cdot\gamma6_{nf}(z-z02)}\right)\cdot FD_{nf}(x)\quad,$$

$$E^{III(7)}_{y}(x,z) = \sum_{nf=-\infty}^{\infty} T7_{nf}e^{-CJ\cdot\gamma7_{nf}\cdot(z-z02)}\cdot FD_{nf}(x)\quad,$$

где $T3_{nf}$ и $R3_{nf}$ – неизвестные комплексные амплитудные коэффициенты прошедшего и отраженного полей в "канале Флоке" конечной длины, $T5_{nf}, T6_{nf}$ и $R5_{nf}, R6_{nf}$ – неизвестные комплексные амплитудные коэффициенты прошедшего и отраженного поля в диэлектрических слоях "канала Флоке"; $T7_{nf}$ – неизвестные комплексные амплитудные коэффициенты внешнего пространства, $FD_{nf}(x)$ – ортонормированные поперечные собственные

функции "канала Флоке", $\gamma 3_{mf}$ – продольные коэффициенты распространения волны в "канале Флоке" конечной длины, $\gamma 5_{mf}$, $\gamma 6_{mf}$ и $\gamma 7_{mf}$ – в диэлектрических слоях "канала Флоке"(внешнего пространства).

После уменьшения количества неизвестных комплексных амплитудных коэффициентов (выражая $E_y^{III(5)}(x,z), E_y^{III(6)}(x,z)$ через $E_y^{III(7)}(x,z)$, как показано в [6]) и применения процедуры Галеркина была получена СЛАУ относительно $T3_{mf}, R3_{mf}, T7_{mf}$ которая была решена методом редукции. После определения указанных коэффициентов, а также компонент возбуждающего поля и сшивания полей областей I и II при z=-z2 было найдено поле отраженных волн.

Результаты работы

На рис. 2 – 6 представлены зависимости модуля коэффициента отражения для волны H_{10} для различных поперечных размеров ФАР с СПС и диэлектрическим заполнением.

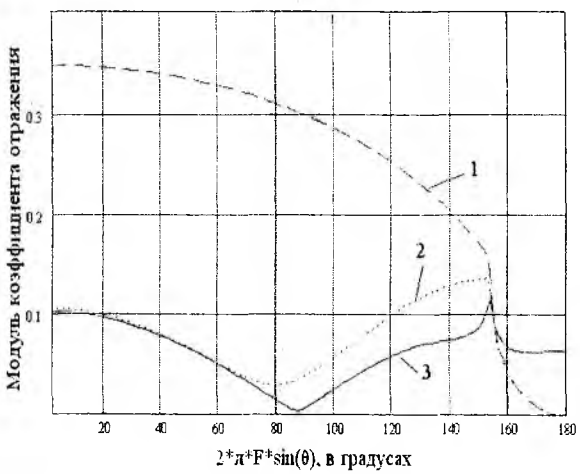


Рис.2. WX=0.5714λ, PX=0.5714λ:1 – z1=0, z2=0 ; 2–z1=0.05λ; z2=0.25λ; 3– z1=0.05λ; z2=0.25λ; z01=0.05λ e5=1.2

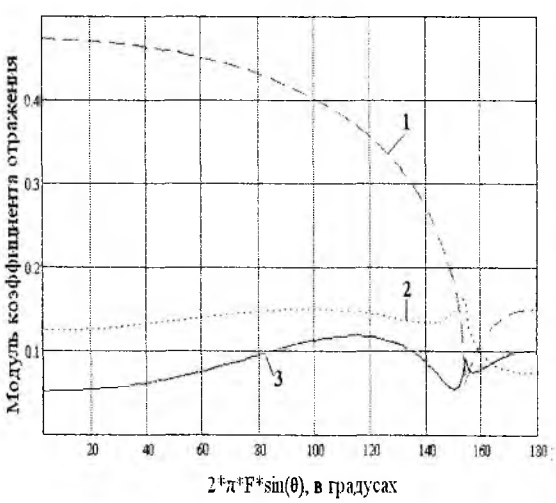


Рис.3. WX=0.5714*0.937λ; PX=0.5714λ: 1– z1=0, z2=0 ; 2 – z1=0.1λ; z2=0.45λ; z21 = =0.55 λ; ε3=1.2; 3 – z1=0.45λ; z2=0.6λ

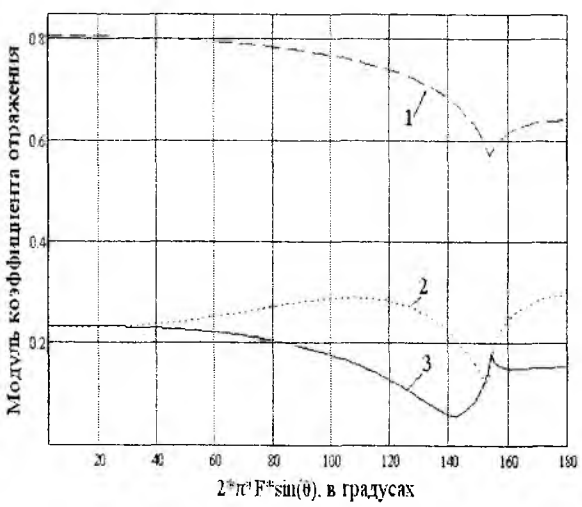


Рис.4. WX=0.5714*0.88λ; PX=0.5714λ: 1– z1=0, z2=0 ; 2 – z1=0.45λ; z2=0.6λ; 3 – z1=0.1λ; z2=0.45λ; z21=0.55 λ; ε3=1.2 .

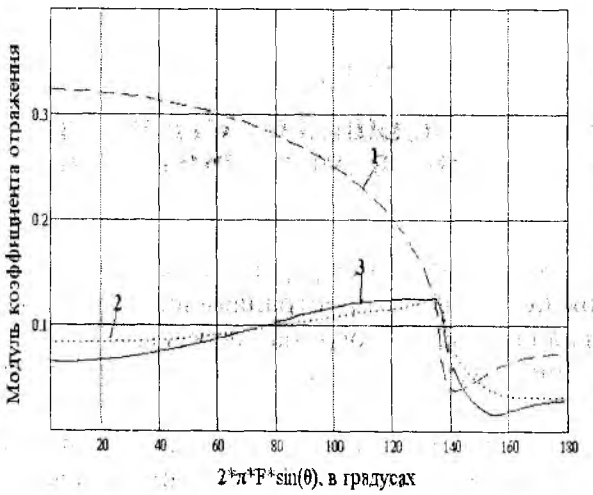


Рис.5. WX=0.937*0.6205λ, PX=0.6205λ:
1– z1=0, z2=0 ; 2 – z1=0.05λ; z2=0.25λ;
3 – z1=0.5λ; z7=0.3λ; ε4=1.7

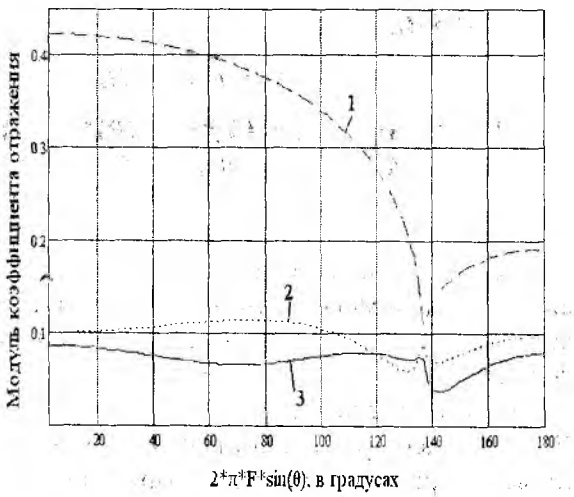


Рис.6. WX=0.88*0.6205λ, PX=0.5714λ:
1– z1=0, z2=0 ; 2 – z1=0.05λ; z2=0.25λ;
3 – z1=0.6λ; z2=0.85λ; ε4=1.4

Для случая ФАР с бесконечными тонкими стенками волноводов: W=F=0.5714 СПС с с диэлектрическим заполнением (z1=0.05λ; z2=0.25λ; z01=0.05λ, e5=1.2) дает незначительное улучшение согласования по сравнению с СПС без диэлектрического заполнения (z1=0.05λ; z2=0.25λ) , соотношение модулей коэффициентов отражения 0.12/0.14.

Для ФАР с поперечными размерами F=0.5714λ и W=0.937F СПС с диэлектрическим заполнением (z1=0.1λ; z2=0.45λ; z21=0.55 λ) не дает уменьшения модуля коэффициента отражения в сравнении с СПС без заполнения (z1=0.45λ; z2=0.6λ). При F=0.6205λ и W=0.937F (рис.5), как видно из кривых зависимости модуля коэффициента отражения для СПС с (z1=0.5λ; z7=0.3λ; ε4=1.7) /без диэлектрического заполнения (z1=0.05λ; z2=0.25λ) различие между ними мало.

Для случая ФАР : F=0.5714λ и W=0.88F (рис.4) СПС с диэлектрическим заполнением (z1=0.1λ; z2=0.45λ; z21=0.55 λ; ε3=1.2) обеспечивает лучшее согласование нежели без заполнения (z1=0.45λ; z2=0.6λ), соотношение модулей коэффициентов отражения (с диэлектриком /без диэлектрика) 0.235/0.31. Для F=0.6205λ и W=0.88F (рис.6) СПС с заполонением (z1=0.6λ; z2=0.85λ; ε4=1.4) и без него (z1=0.05λ; z2=0.25λ), соотношение модулей коэффициентов отражения 0.085/0.115.

Выводы

Численное исследование согласования ФАР с внешним пространством за счет применения согласующей структуры показало, что с увеличением толщины СПС её использование становится более заметным. Применение диэлектрических заполнений в СПС целесообразно применять внутри структуры, поскольку слои покрытия не вносят существенный вклад в уменьшение коэффициента отражения. Уменьшение модуля коэффициента отражения при использовании СПС и диэлектрического заполнения в сравнении с СПС без заполнения составляет 25 – 35%.

Список литературы: 1. *Амитей, Р., Галиндо, В., Ву, Ч.* Теория и анализ фазированных антенных решёток (пер с англ.) – М. : Мир, 1974. – 455 с. 2. *Гостюхин, В.Л., Гринева, К.И., Трусов, В.Н.* Вопросы проектирования активных ФАР с использованием ЭВМ / под ред. В.Л. Гостюхина. – М. : Радио и связь, 1983. – 248 с. 3. *Морозов, В.М., Магро, В.И., Марченко, С.В.* Применение метода интегрального уравнения для решения трехмерных дифракционных задач // Электродинамика и физика СВЧ. – 2006. – С. 56–61. 4. *Марченко, С.В, Морозов, В.М., Съянов, А.М.* Электродинамический расчет волноводных ФАР с магнитодиэлектрическим заполнением методом пронизывающей области // Радиотехника. – 2011. – №167. – С.4-9.

Днепродзержинский государственный технический университет

Поступила в редколлегию 10.09.2012

УДК 551.501.8:621.396.96

В.М. КАРТАШОВ, д-р техн. наук, Д.Н. КУЛЯ

СИНТЕЗ И АНАЛИЗ ДИСКРИМИНАТОРА СЛЕДЯЩЕГО УСТРОЙСТВА СИСТЕМ РАДИОАКУСТИЧЕСКОГО ЗОНДИРОВАНИЯ АТМОСФЕРЫ

Введение

Системы радиоакустического зондирования (РАЗ) атмосферы используют явление рассеяния электромагнитной волны на звуковой посылке, распространяющейся в атмосфере. По характеристикам принимаемого радиосигнала, рассеянного на звуковых колебаниях, определяется скорость звука в среде, позволяющая вычислять далее значения температуры воздуха и скорости ветра [1].

В настоящее время в системах РАЗ реализованы алгоритмы обработки сигналов, заимствованные из радиолокации, которые основаны на выделении и измерении доплеровской частоты колебаний. Однако оценки скорости звука, получаемые в соответствии с используемыми в системах РАЗ алгоритмами, характеризуются наличием значительных по величине систематических погрешностей (по температуре величина смещения достигает единиц градусов) [1 – 3].

Цель работы – повышение точности оценок скорости звука, получаемых в системах РАЗ.

Основная задача статьи – синтез упрощенного оптимального по критерию максимума правдоподобия алгоритма оценивания параметров радиосигнала, рассеянного акустическим волновым пакетом.

Принцип работы систем радиоакустического зондирования атмосферы

Системы РАЗ – это системы, предназначенные преимущественно для измерения температуры атмосферы. Принцип работы таких систем состоит в том, что в атмосферу излучаются акустический сигнал и радиосигнал соответственно акустической и радиоантенной.

Акустическая посылка создает в атмосфере систему неоднородностей. Электромагнитный сигнал отражается от системы неоднородностей, при этом изменяется его структура. По отраженному радиосигналу, принимаемому второй радиоантенной, оценивается скорость звука и вычисляются неизвестные параметры атмосферы по формуле (1)

$$c_S = (\gamma RT/\mu)^{1/2} + \vec{V}\vec{n} = a_\kappa \sqrt{T_0} + \vec{V}\vec{n}, \qquad (1)$$

где γ – показатель адиабаты (отношение теплоемкостей при постоянном давлении C_p и постоянном объеме C_v); R – универсальная газовая постоянная; T_0 – абсолютная температура; μ – молекулярный вес воздуха; $\vec{V}$ – вектор скорости ветра; $\vec{n} = \vec{k}_s / |\vec{k}_s|$; $\vec{k}_s = 2\pi/\lambda_s$ – волновой вектор. Коэффициент $a_\kappa = (\gamma R/\mu)^{1/2}$ для воздуха слабо зависит от давления, температуры, влажности и на практике обычно считают $a_\kappa = const$.

Особенности рассеяния электромагнитного сигнала на звуке

Рассеяние радиосигнала на акустической волновой посылке имеет некоторые особенности: во-первых, рассеяние радиоволны на звуке является частотно-зависимым: уровень рассеянного радиосигнала зависит от значения параметра расстройки условия Брэгга $q = 2k_e - k_S$, где k_e – волновое число электромагнитного колебания, k_S – волновое число акустического колебания, а во-вторых, при рассеянии на звуке существенно изменяется структура излучаемого радиосигнала, который приобретает в процессе рассеяния дополнительную амплитудную и угловую модуляцию, что видно на рис. 1, на котором изображен спектр радиосигнала, отраженного от акустического пакета на разных высотах. Спектр является не симметричным, что имеет место при совместной амплитудно-фазовой модуляции колебаний.

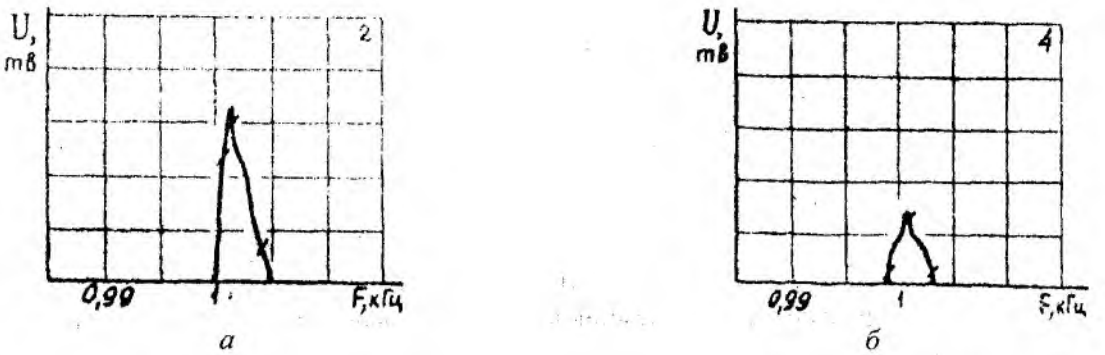


Рис. 1. Спектр радиосигнала, отраженного от акустического пакета на разных высотах: а – 64 *м*, б – 90 *м*

Методы обработки сигналов

Рассмотрим существующие методы обработки сигналов в радиоакустических системах. Для этого рассмотрим рис. 2, на котором приведены три профиля, полученные с использованием классических корреляционных алгоритмов обработки при трех различных значениях звуковой частоты. Кружками на профилях отмечены те высоты, на которых f_0 практически совпадала с f_s, т.е. высоты для которых $q = 0$. Именно на этих высотах значение оцениваемой температуры оказалось наиболее точным.

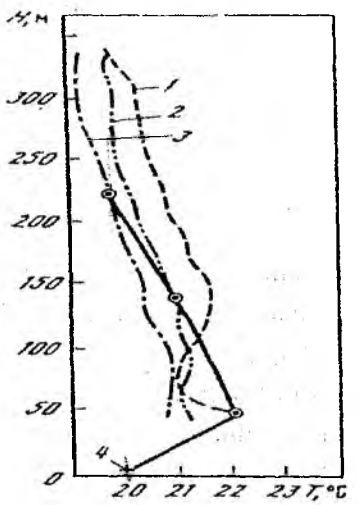


Рис. 2. Профили температуры в атмосфере

Некоторые применяемые на практике методы направлены на поиск условия, когда q в точке будет равно нулю. Такой поиск обеспечивается за счет изменения частоты звукового сигнала, потому методы называются методами с подстройкой частоты зондирующих сигналов. При изменении значения частоты звукового излучения и неизменном значении частоты радиосигнала изменяются значения параметра $q = 2k_e - k_s$. При этом точке $q = 0$ соответствует максимум амплитуды отраженного сигнала и в точке $q = 0$ доплеровский сдвиг частоты рассеянного радиосигнала совпадает с частотой излучаемой звуковой посылки.

Методы с подстройкой частоты зондирующих сигналов требуют большого количества времени для зондирования. Поэтому в ряде систем зондирования применяется метод без подстройки частоты зондирующих колебаний, основанный на оценке доплеровской частоты принимаемого радиосигнала, которая производится путем классической многоканальной корреляционной обработки, показанной на рис. 3, где в качестве опорных сигналов корреляторов применяются колебания той же формы, что и излученные, но с разными несущими частотами.

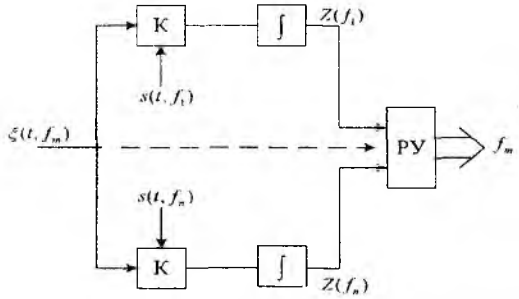


Рис. 3. Корреляционный приемник для оценки доплеровского сдвига частоты

Принцип работы алгоритма описывается выражением

$$Z(f_i, f_m) = \frac{2}{N}\int_0^T \xi(t, f_m)\cdot s(t, f_n)dt, \quad i = \overline{1,n}, \tag{2}$$

где $\xi(t, f_m)$ – принятый электромагнитный сигнал со значением несущей частоты f_m.

Однако такой алгоритм характеризуется систематическими погрешностями в определении температуры воздуха и необходимостью использовать многоканальный коррелятор.

Возникновение погрешности

Чтобы разобраться в причинах возникновения погрешностей, рассмотрим апостериорное распределение информативного параметра a принимаемого радиосигнала [4]

$$P(a|y) = c'' P(a)\exp\left[-\frac{1}{N_0}\int_0^T y(t)u_c(t,a)dt\right]\exp\left[-\frac{E_a}{N_0}\right], \tag{3}$$

где c'' – константа, определяющая масштаб по оси ординат; $E_a = \int_0^T u_c^2(t,a)dt$ – энергия сигнала при данном значении сообщения a; $y(t)$ – смесь полезного сигнала и помехи на входе устройства обработки и воспроизведения информативного параметра a; $u_c(t,a)$ – сигнал произвольной формы, известный в месте приема точно, за исключением неизвестного информативного параметра a с априорным распределением $P(a)$.

Для широкого класса локационных задач (в том числе задач измерения скорости объектов по доплеровской частоте) и используемых сигналов полагается, что форма сигнала $y(t)$ не отличается от формы опорного сигнала $u_c(t,x)$, а величина E_a полагается не зависящей от информативного параметра a, в процессе формирования искомой оценки она не учитывается и включается в константу c''. Однако в системах РАЗ атмосферы указанное условие не выполняется. Во-первых, рассеяние радиоволны на звуке является частотно-зависимым: уровень рассеянного радиосигнала зависит от скорости звука, которую необходимо оценить, а во-вторых, при рассеянии на звуке существенно изменяется структура излучаемого радиосигнала, который приобретает дополнительную амплитудную и угловую модуляцию [1].

Следящая система РАЗ атмосферы

Для устранения систематических погрешностей необходимо учитывать изменение формы радиосигнала при его взаимодействии с акустическим колебанием. Форма рассеянного радиосигнала может быть определена с использованием функции рассеяния (4) [2, 3]:

$$F(r,q) = \int_{-\infty}^{\infty} E(2r'-r)S^*(r')e^{jqr'}dr', \tag{4}$$

где $q = 2k_e - k_s$ – параметр расстройки условия Брэгга; $k_s = 2\pi f_a / c_a$ – волновое число для звука; $k_e = 2\pi f / c$ – волновое число электромагнитной волны; r – смещение сигналов по

координате «дальность»; $E(2r'-r)$ – комплексная пространственная огибающая радиосигнала; $S(r')$ – комплексная пространственная огибающая акустического сигнала.

В качестве информационного параметра необходимо использовать расстройку условия Брэгга q, которая функционально связана со скоростью звука (5).

$$c_S = \frac{2\pi \cdot f_S}{4\pi \cdot f/c - q}. \tag{5}$$

При формировании опорных сигналов необходимо учитывать зависимость энергии формируемых сигналов от q.

Для упрощения схем многоканальной корреляционной обработки, используемых в существующих системах РАЗ атмосферы, в радиоакустических системах необходимо реализовать следящий алгоритм, который представляет собой упрощенный оптимальный алгоритм обработки сигнала, где упрощение достигается за счет использования значения оцениваемого параметра, определенного на предыдущих циклах наблюдения. Следящий алгоритм соответствует уравнениям:

$$\left(\int_0^T F(r,q_0)F(r,q_1)dr - E(q_1)\right) - \left(\int_0^T F(r,q_0)F(r,q_2)dr - E(q_2)\right) = 0, \tag{6}$$

$$\left(\int_0^T F(r,q)F(r,q_1)dr - E(q_1)\right) - \left(\int_0^T F(r,q)F(r,q_2)dr - E(q_2)\right) = x, \tag{7}$$

где $F(r,q_1)$, $F(r,q_2)$ – опорные сигналы, сформированные с использованием функции рассеяния; $F(r,q_0), F(r,q)$ – сигналы с соседних циклов измерения; x – разница, по которой определяется величина отличия между q_0 и q; $E(q_1)$, $E(q_2)$ – корректирующие слагаемые, соответствующие энергетическому уровню опорных сигналов.

Схема следящего устройства обработки радиосигнала на основе двухканального коррелятора, устраняющая систематическую ошибку показана на рис. 4 [5].

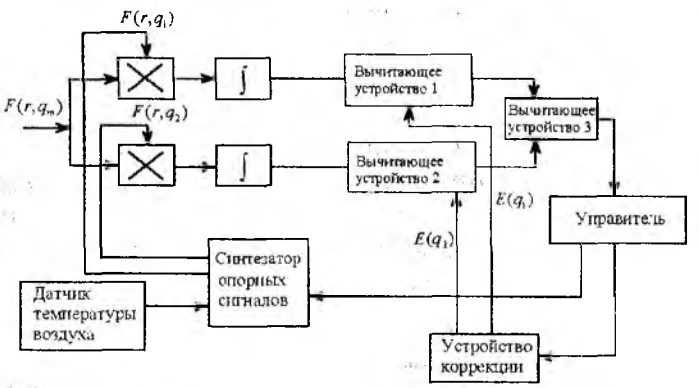


Рис. 4. Следящее устройство обработки сигналов систем РАЗ атмосферы

Схема состоит из датчика температуры, дискриминатора, управителя, генератора опорных сигналов и устройства коррекции. Датчик температуры измеряет температуру на первом уровне трассы зондирования и в соответствии с измеренной температурой формирует управляющий сигнал. Сигнал с датчика температуры задает генератору опорных сигналов значения параметров расстройки условия Брэгга q_1 и q_2, с которыми необходимо сформировать опорные сигналы для второго уровня трассы зондирования, такие, что разница между корреляционными интегралами сигнала принимаемого с первого уровня трассы зондирования и опорных сигналов равнялась нулю (6).

Опорные сигналы поступают на устройство коррекции, которое формирует корректирующий сигнал в соответствии с энергией опорных сигналов и на дискриминатор, состоя-

щий из двухканального корреляторa и трех вычитающих устройств. Дискриминатор вычисляет разницу U (8) между корреляционными интегралами сигнала, принимаемого с некоторого уровня трассы зондирования, и опорных сигналов удовлетворяющих уравнению (6):

$$\left(\int_0^T F(r,q)F(r,q_1)dr - E(q_1)\right) - \left(\int_0^T F(r,q)F(r,q_2)dr - E(q_2)\right) = U \qquad (8)$$

где $F(r,q)$ – сигнал, принимаемый с некоторого уровня трассы зондирования.

Дискриминатор определяется дискриминационной характеристикой (ДХ). Дискриминационная характеристика – это характеристика, показывающая зависимость напряжения U на выходе дискриминатора от разницы $y = q - q_0$, где q – расстройка условия Брэгга на некотором уровне трассы зондирования, q_0 – расстройка условия Брэгга на предыдущем уровне трассы зондирования. Наибольшая линейность и длина линейного участка ДХ для заданных зондирующих колебаний задаются величиной 2Δ, на которую отличаются параметры расстройки условия Брэгга q_1 и q_2 опорных сигналов.

Напряжение с дискриминатора поступает на управитель, который в соответствии с дискриминационной характеристикой дискриминатора вычисляет расстройку условия Брэгга на текущем уровне трассы зондирования и формирует управляющий сигнал, который задает генератору опорных сигналов значения параметров расстройки условия Брэгга, при которых напряжение на выходе дискриминатора для сигнала с текущего уровня трассы зондирования будет стремиться к нулю.

Заключение

Синтезирован упрощенный оптимальный алгоритм обработки сигналов и сформулированы предложения по совершенствованию существующих систем РАЗ. В результате синтеза получен дискриминатор, позволяющий производить оценки без систематической ошибки, свойственной системам РАЗ, построенным как измерители доплеровской частоты. Устранение систематической ошибки достигается за счет формирования опорных сигналов дискриминатора в соответствии с преобразованиями, которым сигнал подвергается в локационном акустическом канале. Опорные сигналы формируются с помощью функции рассеяния. Значения корреляционных интегралов в дискриминаторе корректируются с учетом энергий формируемых опорных сигналов.

Скорость звука в предложенном алгоритме оценивается по энергетическому параметру сигнала, в связи с чем, значения корреляционных интегралов в двухканальной схеме корректируются с учетом энергий формируемых опорных сигналов. Оценка скорости звука по энергетическому параметру достаточно непривычна, поскольку в радиолокации утвердилась точка зрения, что определение скорости движения объекта осуществляется посредством оценки неэнергетических параметров сигнала, чаще всего частоты колебания.

Список литературы 1. *Каллистратова, М.А., Кон, А.И.* Радиоакустическое зондирование атмосферы. – М. : Наука, 1985. – 200 с. 2. *Карташов, В.М.* Функции рассеяния сигналов систем зондирования атмосферы // Радиотехника. – 2001. – №118. – С. 61-65. 3. *Карташов, В.М.* Модели и методы обработки сигналов систем радиоакустического и акустического зондирования атмосферы. – Харьков : ХНУРЭ, 2011. – 233 с. 4. *Теоретические* основы радиолокации / Я.Д. Ширман, В.Н. Голиков, И.Н. Бусыгин и др. ; под ред. Я.Д. Ширмана. – М. : Сов. радио, 1970. – 560 с. 5. *Карташов, В.М., Куля, Д.Н., Пащенко, С.В.* Алгоритм автосопровождения изменений информационного параметра сигнала радиоакустических систем // Восточно-Европейский ЖУРНАЛ передовых технологий. – 2012 – № 4/9(58). – С. 57-61.

Харьковский национальный университет радиоэлектроники

Поступила в редколлегию 13.09.2012

УДК: 681.128.82

Б.В. ЖУКОВ, канд. техн. наук, А.В. ОДНОВОЛ

КОНТРОЛЬ УРОВНЯ ЛЕГКОИСПАРЯЮЩИХСЯ ЖИДКОСТЕЙ МЕТОДОМ АКУСТИЧЕСКОЙ ЛОКАЦИИ

Введение

Измерение и текущий контроль уровня жидких, в том числе легкоиспаряющихся сред, по-прежнему, представляет актуальную задачу в процессе их производства, хранения и отпуска. Для этих целей применяются различные методы измерения уровня, в том числе и метод акустической локации.

Уровнемеры [1 – 3], в которых используется плоская акустическая волна, обеспечивают разрешение уровней жидкостей $\pm(0{,}2...1)$ *мм*, что определяет перспективность их использования не только для технологических, но и для коммерческих целей.

В работе рассмотрены особенности функционирования акустических уровнемеров с абсолютным и относительным методами измерения в процессе контроля уровня нескольких типов жидкостей, включая воду, дизельное топливо, спирт, растворитель и автомобильные бензины. Результаты исследований, выполненных на рабочих частотах 2 и 10 *кГц*, позволили определить области применения акустических уровнемеров для контроля уровня жидких сред.

Методика проведения исследований

Известно, что скорость распространения акустической волны зависит от различных параметров, в том числе температуры [4], давления и насыщенности паров контролируемой жидкости в канале распространения над ее поверхностью. При абсолютном методе измерения уровня влияние температуры компенсируется введением в состав уровнемера датчика температуры, а влиянием давления и насыщенности паров контролируемой жидкости в большинстве случаев пренебрегают. Относительный метод, за счет введения эталонного канала, теоретически позволяет компенсировать влияние изменений всех перечисленных факторов. Поэтому относительный метод считается предпочтительным при существенных изменениях параметров канала распространения, наблюдающихся, например при контроле уровня легкоиспаряющихся жидкостей.

Локационный метод определения уровня заключается в измерении расстояния до поверхности контролируемой жидкости и последующего расчета уровня по известным параметрам резервуара. Поэтому исследование влияния паров над контролируемой поверхностью жидкости, то есть в самом канале распространения акустической волны, целесообразно проводить в режиме измерения расстояния. При этом в течение проведения исследований необходимо обеспечивать неизменность геометрического расстояния «плоскость отсчета уровнемера – поверхность жидкости», а также температуры и давления окружающей среды.

В уровнемерах с плоской акустической волной [1] измерение расстояния до контролируемой поверхности производится в звуковедущей трубе (звуководе). Поэтому следует ожидать, что сразу после установки звуковода в контролируемую жидкость в нем начинается процесс возрастания насыщенности паров над поверхностью жидкости. Этот процесс будет продолжаться до достижения внутри звуковода динамического равновесия «испарение – конденсация». Если скорость распространения акустической волны в парах контролируемой жидкости отличается от ее величины в воздушной среде, то в интервале времени от момента установки звуковода до достижения динамического равновесия уровнемер будет регистрировать некоторое изменение расстояния до контролируемой поверхности. Следовательно, непосредственно после установки звуковода показания индикатора уровнемера будут соответствовать геометрическому расстоянию до контролируемой поверхности, а после наступления динамического равновесия паров внутри звуковода они могут отличаться от исходных.

При этом отличие от геометрического расстояния будет определяться как концентрацией паров контролируемой жидкости, так и скоростью акустической волны в них.

Рассмотренная ситуация будет наблюдаться при использовании абсолютного метода измерения расстояния до поверхности контролируемой жидкости. При относительном методе между первым (верхним) и вторым (нижним) приемниками внутри звуковода формируется участок известной (эталонной) протяженности [1]. Если эталонный канал расположен внутри резервуара и его протяженность соизмерима с протяженностью измерительного канала "второй приемник – поверхность контролируемой жидкости", то можно ожидать совпадения величин скоростей акустической волны в обоих каналах. Поэтому применение относительного метода должно существенно или полностью устранять влияние наличия паров контролируемой жидкости на скорость распространения акустической волны, по крайней мере, после наступления динамического равновесия в звуководе. Кроме того, уровнемер с относительным методом измерения [1] обеспечивает возможность раздельного текущего контроля времен прохождения акустического эталонного и измерительного каналов, что позволяет исследовать динамику движения паров контролируемых жидкостей вдоль звуковода.

Таким образом, в процессе исследований влияния паров контролируемых жидкостей на функционирование акустических уровнемеров необходимо:

1.Обеспечивать неизменность геометрического расстояния «плоскость отсчета уровнемера – поверхность контролируемой жидкости»;

2.Регистрировать результаты измерения уровнемером расстояния до поверхности контролируемой жидкости с момента установки звуковода в контролируемую жидкость и до достижения внутри звуковода динамического равновесия «испарение – конденсация» в случае использования прибора с абсолютным методом. При относительном методе также регистрировать время прохождения акустической волной эталонного канала.

Оценка влияния паров контролируемых жидкостей при использовании уровнемера с абсолютным методом будет проводиться по величине текущего изменения расстояния

$$\Delta R(t) = R_i(t) - R_0,$$

где $R_i(t)$ – текущее измеряемое расстояние до поверхности жидкости, а R_0 - расстояние в момент установки звуковода в контролируемую жидкость.

В случае использования уровнемера с относительным методом будут также оцениваться величины изменения времени прохождения акустической волной эталонного и измеряемого расстояний

$$\Delta T_{\acute{y}}(t) = T_{\acute{y}i}(t) - T_{\acute{y}0},$$

где $T_{\acute{y}_i}(t)$ – текущее время прохождения акустической волной эталонного канала, а $T_{\acute{y}0}$ - время прохождения эталонного канала в момент установки звуковода в контролируемую жидкость, и

$$\Delta T_{\tilde{O}}(t) = T_{\tilde{O}i}(t) - T_{X0}.$$

Здесь $T_{\tilde{O}i}(t)$ - текущее время прохождения акустической волной расстояния от нижнего приемника до контролируемой поверхности, а T_{X0} - время прохождения этого же расстояния в момент установки звуковода в контролируемую жидкость.

Исследования проводились на рабочих частотах 2 и 10 *кГц* как для слабо (вода, дизельное топливо), так и легко испаряющихся (спирт, растворитель 650 и несколько типов автомобильных бензинов) жидкостей при различных длинах и диаметрах звуководов.

Результаты исследований

На рис. 1 представлены результаты контроля изменения расстояния $\Delta R(t)$ до поверхностей воды (кривая 1), дизельного топлива (кривая 2), растворителя 650 (кривая 3), спирта (кривая 4) и автомобильных бензинов АИ-98 (кривая 5) и АИ-95 «prim» (кривая 6) в течение первых 90 *мин* после установки в них звуковода уровнемера с абсолютным методом измерения.

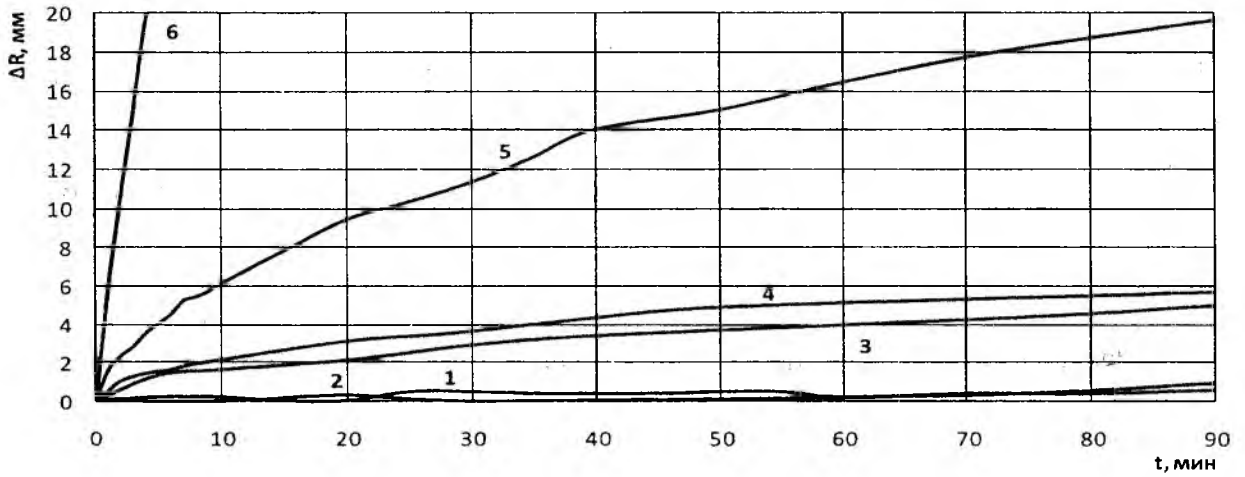


Рис. 1

Приведенные результаты свидетельствуют, что для слабо испаряющихся жидкостей (вода и дизельное топливо) величина $\Delta R(t)$ не превышала 1 *мм* в течение всего времени наблюдения. Для легкоиспаряющихся жидкостей (растворитель, спирт и бензины) на индикаторе уровнемера регистрировалось возрастание расстояния до поверхности контролируемых жидкостей. Этот процесс обычно продолжался в течение (1,5…2) *ч* до установки в звуководе динамического равновесия. Последующие изменения $\Delta R(t)$ не превышали ± 2 *мм* в течение длительного времени (см. рис.2). На этом рисунке представлена временная зависимость $\Delta R(t)$ для бензина АИ-98 через несколько часов (начало отсчета времени на рис.2) после наступления динамического равновесия паров бензина в звуководе.

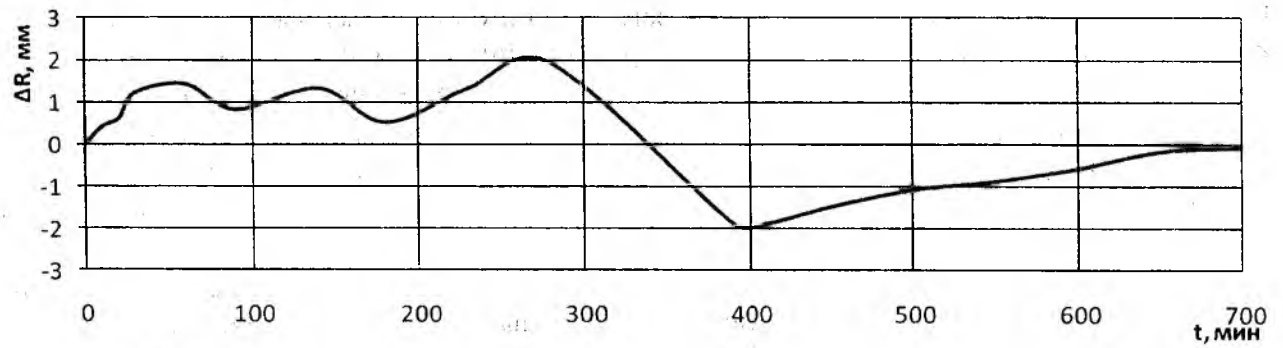


Рис.2

Дальнейшие исследования, учитывая практическую актуальность, проводились для автомобильных бензинов, произведенных из газового конденсата (прямогонный, АИ-95 и АИ-98) и нефти (АИ-92, АИ-95 и АИ-95«prim»). Например, на рис.3 представлены экспериментальные величины $\Delta R(t)$ для АИ-98, измеренные через 50 *мин* после установки звуковода в бензин. Они показывают (см. табл. 1), что для рабочих частот 2 и 10 *кГц* в диапазонах длин звуководов от 930 до 3130 *мм*, диаметров звуководов от 22 до 88 *мм* и температур окружающей среды от 12,5 до 15^{0}С величины $\Delta R(t)$ составляли (12 ± 1) *мм*.

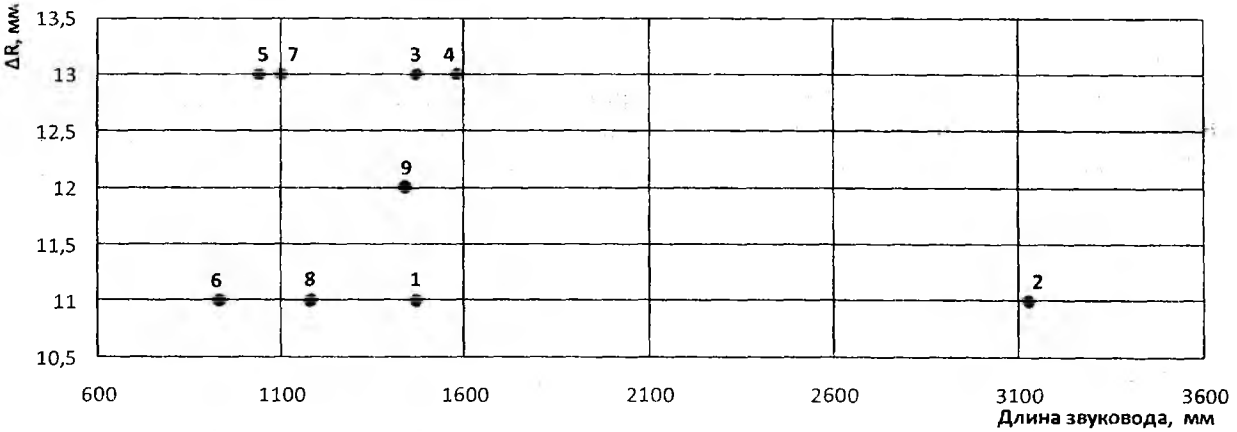


Рис.3

Таблица 1

Номер точки	Диаметр звуковода, *мм*	Длина звуковода, *мм*	Температура, градусы	Рабочая частота, *кГц*	Величина $\Delta R(t)$, *мм*
1	22	1470	13	2	11
2	22	3130	15	2	11
3	27	1470	13	10	13
4	27	1580	13	2	13
5	36	1040	14	2	13
6	36	930	14	10	11
7	48	1100	14	10	13
8	48	1180	12.5	2	11
9	88	1440	13	2	12

Проведенные исследования свидетельствовали, что для каждого типа бензина величина $\Delta R(t)$ практически не зависела от длины и диаметра звуковода (см. рис.3) и существенно зависела от исходного сырья и технологии их производства (см. кривые 5 и 6 на рис.1). Последнее обстоятельство очевидным образом связано с существенным отличием химического состава бензинов, а значит и их паров, произведенных их различного исходного сырья. На рис.4 приведены зависимости $\Delta R(t)$ для бензина АИ-98 (кривая 1), произведенного из газового конденсата, и бензина АИ-95«prim» (кривые 2 и 3), произведенного из нефти.

Данные на рис.4 позволяют уменьшить расстояние до поверхности бензина с 900 *мм* (кривая 2) до 500 *мм* (кривая 1), при котором наблюдается отсутствие зависимости величины $\Delta R(t)$ от длины звуковода.

С целью снижения влияния паров бензинов на результаты контроля расстояния до контролируемой поверхности был применен относительный метод измерения. На рис. 5 представлены временные зависимости $\Delta R(t)$ для бензина АИ-92 при абсолютном (кривая 1) и относительном (кривая 2) методах измерения. Представленные данные свидетельствуют, что применение относительного метода не устраняет, а обеспечивает только незначительное снижение влияния паров бензина на результаты измерения расстояния до контролируемой поверхности.

На рис. 6 представлены результаты измерения увеличения времен прохождения акустическим сигналом измеряемого $\Delta T_X(t)$ (кривая *1*) и эталонного $\Delta T_Y(t)$ (кривая *2*) расстояний для бензина марки АИ-98 при относительном методе. Они свидетельствуют, что несмотря на существенное увеличение времени прохождения измеряемого расстояния $\Delta T_X(t)$, время

прохождения эталонного расстояния $\Delta T_{\hat{y}}(t)$ увеличивается незначительно, оставаясь практически неизменным. Данный факт позволяет сделать вывод о том, что пары бензина, существенно снижающие скорость распространения акустической волны, в основном локализуются непосредственно над его поверхностью и лишь их незначительная часть поднимается вверх по звуководу. Именно по этой причине относительный метод также не может считаться пригодным для контроля уровня автомобильных бензинов без применения специальных мер, устраняющих их испарение.

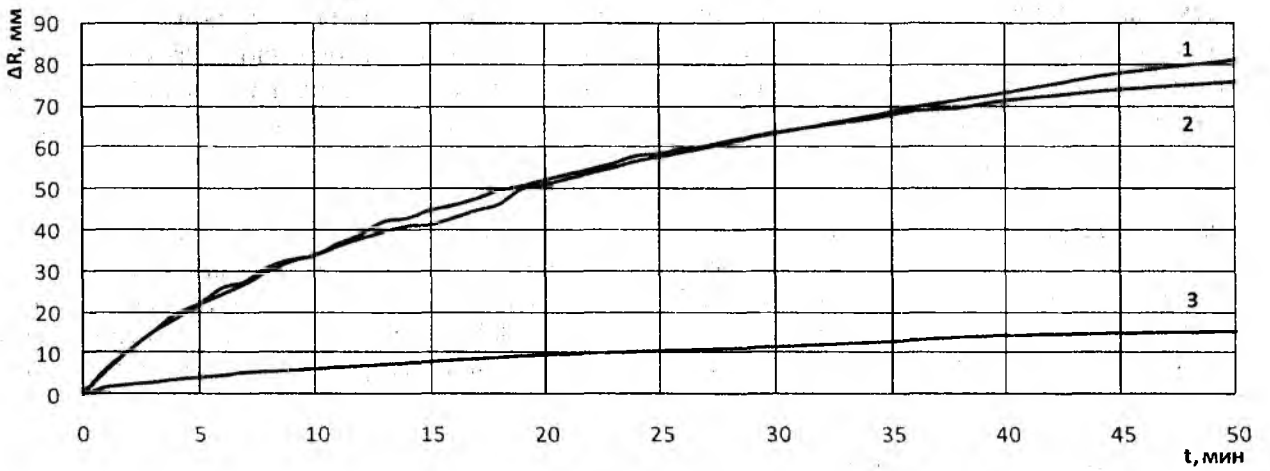


Рис.4

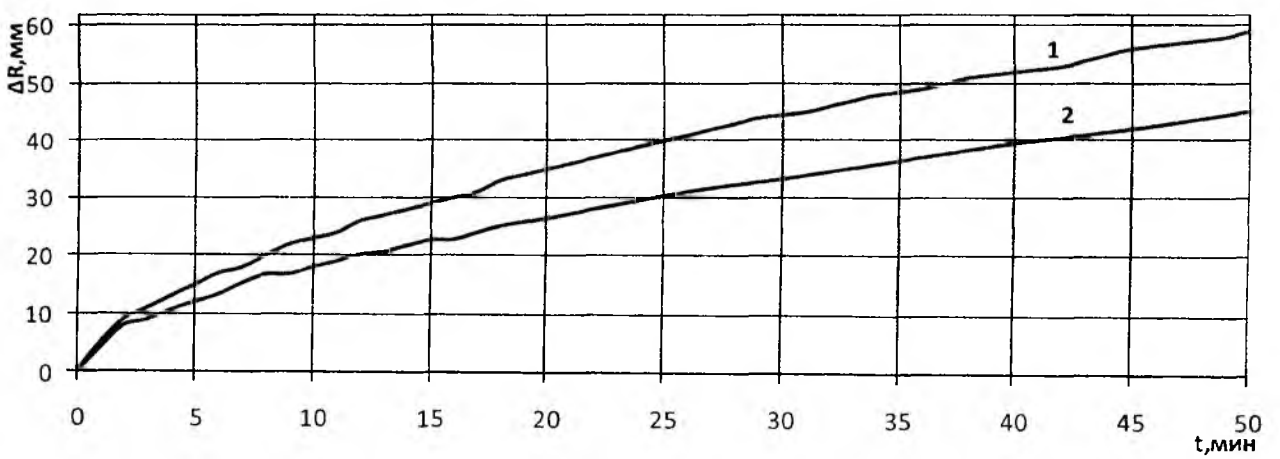


Рис.5

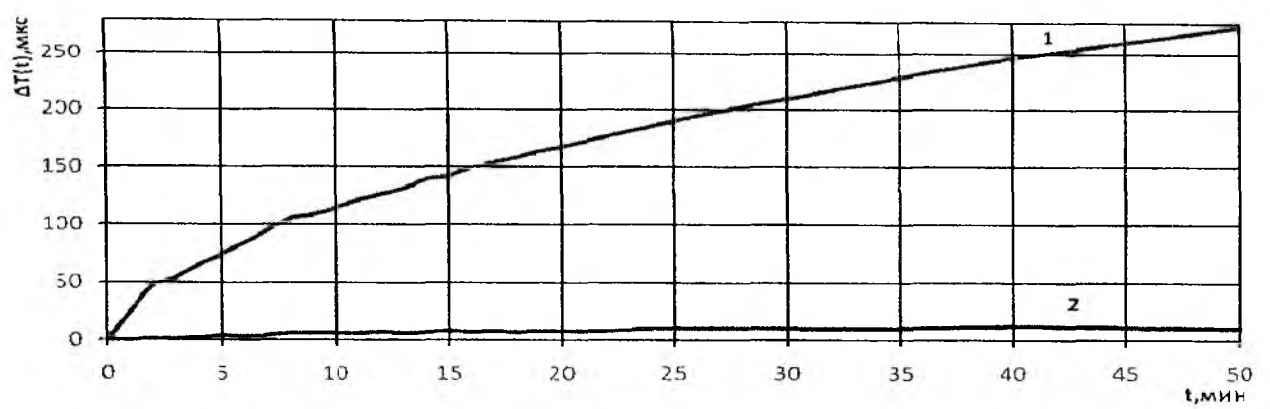


Рис.6

Выводы

Результаты исследований свидетельствуют:

- метод акустической локации целесообразно применять для контроля уровня слабо испаряющихся жидкостей (вода, дизельное топливо и др.);

- пары легкоиспаряющихся жидкостей (бензин, спирт, растворитель), в которых скорость распространения акустической волны может быть существенно ниже по сравнению с воздушной средой, локализуются непосредственно над поверхностью, что исключает перспективность использования относительного метода для контроля их уровня;

- максимальная концентрация паров автомобильных бензинов формируется в слое высотой до 50 *см*, при этом погрешность определения их уровня может превышать 80 *мм* и зависит от химического состава бензина, определяемого типом исходного сырья и технологией его производства.

Список литературы: 1.*Жуков, Б.В., Солярский, Н.Ф.* и др. Низкочастотные модификации акустических преобразователей акустического уровнемера "ЗОНД-3М" // Датчики и системы. – 2007. – №.3. – С.42 – 46. 2. *Пат. Украины* на пол. модель. №2552 G01F23/28. Устройство для измерения уровня жидкости // Бюл. – 2004. – №6. 3.*Пат. РФ* 2146358 G01F 23/296. Устройство для измерения уровня жидких сред // Бюл. – 2000. – №7. 4. *Ржевкин, С.Н.* Курс лекций по теории звука. – М. : МГУ, 1960. – 336с.

Институт радиофизики и электроники
НАН Украины

Поступила в редколлегию 22.09.2012

УДК 535.+543.47+621.396

Г. М. ЧЕКАЛИН, Г. Н. ЧЕКАЛИНА

ОБ АНАЛИТИЧЕСКИХ МЕТОДАХ СИНТЕЗА ПОЛЯРИЗАЦИОННОГО ЭЛЛИПСА

Введение

В работе [1] были получены аналитические выражения для определения компонент поля и разности фаз между ними по заданным значениям угла ориентации β и коэффициента эллиптичности r поляризационного эллипса при повороте системы координат на угол $-\beta$.

В работе [2] получены те же формулы для определения амплитуд компонент поля и разности фаз между ними, но при этом осуществлен поворот системы координат на угол $+\beta$, что вносит определенную путаницу.

В связи с этим возникла необходимость системного анализа данного вопроса.

Задача состоит в том, чтобы исследовать влияние исходной ориентации эллипса поляризации электромагнитной волны на вид аналитических выражений компонент поля, их разности фаз, коэффициента эллиптичности r и угла ориентации β при повороте системы отсчета на угол $\pm\beta$.

Цель работы – показать процедуру определения параметров поляризационного эллипса посредством компонент поля и наоборот, а также уточнить некоторые моменты, связанные с изменениями выражений для определения параметров поляризационного эллипса в связи с преобразованием системы координат.

Решение задачи

Параметрическое уравнение траектории конца вектора $\bar{E}$ в фиксированной точке пространства будет иметь вид

$$\begin{aligned} x &= E_x cos\left(\omega t + \Phi_x\right) \\ y &= E_y cos\left(\omega t + \Phi_y\right). \end{aligned} \tag{1}$$

Годографом параметрической системой (1) является эллипс, уравнение которого имеет вид

$$\frac{x^2}{E_x^2} - 2\frac{xy}{E_x E_y} cos\Phi + \frac{y^2}{E_y^2} = sin^2\Phi. \tag{2}$$

Каноническое параметрическое уравнение эллипса с центром в начале координат и осями, совпадающими с осями координат будет [1]

$$\begin{aligned} x &= a\sin\omega t \\ y &= ra\cos\omega t. \end{aligned} \tag{3}$$

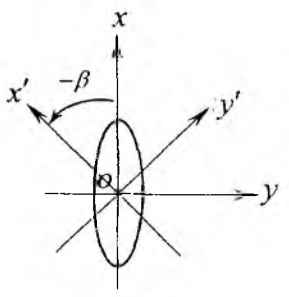


Рис. 1

Повернем систему координат на угол $-\beta$ (рис.1), получим систему преобразования новых координат через старые

$$\begin{aligned} x' &= x cos\beta - y\, sin\beta \\ y' &= x sin\beta + y cos\beta. \end{aligned} \tag{4}$$

Подставим (3) в (4)

$$\begin{aligned} x' &= a cos\beta\, sin\omega t - ra\, sin\beta\, cos\omega t \\ y' &= a sin\beta\, sin\omega t + ra\, cos\beta\, cos\omega t. \end{aligned} \tag{5}$$

Разделим и умножим правую часть первого уравнения системы (5) на $\sqrt{cos^2\beta + r^2 sin^2\beta}$, а правую часть второго уравнения – на $\sqrt{sin^2\beta + r^2 cos^2\beta}$:

$$x' = a\sqrt{cos^2\beta + r^2 sin^2\beta}\left(\frac{-rsin\beta}{\sqrt{cos^2\beta + r^2 sin^2\beta}}cos\omega t - \frac{-cos\beta}{\sqrt{cos^2\beta + r^2 sin^2\beta}}sin\omega t\right),$$
$$y' = a\sqrt{sin^2\beta + r^2 cos^2\beta}\left(\frac{rcos\beta}{\sqrt{sin^2\beta + r^2 cos^2\beta}}cos\omega t - \frac{-sin\beta}{\sqrt{sin^2\beta + r^2 cos^2\beta}}sin\omega t\right). \quad (6)$$

Представим теперь колебания в новой системе координат в виде (1):

$$x' = E'_x\left(cos\Phi'_x cos\omega t - sin\Phi'_x sin\omega t\right) = E'_x cos\left(\omega t + \Phi'_x\right)$$
$$y' = E'_y\left(cos\Phi'_y cos\omega t - sin\Phi'_y sin\omega t\right) = E'_y cos\left(\omega t + \Phi'_y\right), \quad (7)$$

где E'_x, E'_y - амплитуды компонент поля в новой системе координат,

$$E'_x = a\sqrt{cos^2\beta + r^2 sin^2\beta},$$
$$E'_y = a\sqrt{sin^2\beta + r^2 cos^2\beta}, \quad (8)$$

$$cos\Phi'_x = \frac{-rsin\beta}{\sqrt{cos^2\beta + r^2 sin^2\beta}}, \quad sin\Phi'_x = \frac{-cos\beta}{\sqrt{cos^2\beta + r^2 sin^2\beta}}, \quad (9)$$

$$cos\Phi'_y = \frac{rcos\beta}{\sqrt{sin^2\beta + r^2 cos^2\beta}}, \quad sin\Phi'_y = \frac{-sin\beta}{\sqrt{sin^2\beta + r^2 cos^2\beta}}. \quad (10)$$

Исходя из выражений (9), (10), получим

$$tg\Phi'_x = \frac{-cos\beta}{-rsin\beta} = \frac{1}{r}ctg\beta, \quad (11)$$

$$tg\Phi'_y = \frac{-sin\beta}{rcos\beta} = -\frac{1}{r}tg\beta, \quad (12)$$

$$tg\left(\Phi'_y - \Phi'_x\right) = \frac{tg\Phi'_y - tg\Phi'_x}{1 + tg\Phi'_y \cdot tg\Phi'_x},$$

$$tg\left(\Phi'_y - \Phi'_x\right) = \frac{-\frac{1}{r}tg\beta - \frac{1}{r}ctg\beta}{1 + \left(-\frac{1}{r}\right)tg\beta \cdot \frac{1}{r}ctg\beta} = \frac{-\frac{1}{r}\left(tg\beta + ctg\beta\right)}{\left(1 - \frac{1}{r^2}\right)} =$$

$$= \frac{-\frac{1}{r}\left(\frac{sin\beta}{cos\beta} + \frac{cos\beta}{sin\beta}\right)}{\left(1 - \frac{1}{r^2}\right)} = \frac{-\frac{1}{r}\left(\frac{sin^2\beta + cos^2\beta}{sin\beta cos\beta}\right)}{\left(1 - \frac{1}{r^2}\right)} =$$

$$= -\frac{1}{r}\frac{1}{s\sin\beta cos\beta} \div \left(\frac{r^2-1}{r^2}\right) = -\frac{r}{\left(r^2-1\right)s\sin\beta cos\beta}.$$

Помножим числитель и знаменатель правой части уравнения на два:

$$tg\left(\Phi'_y - \Phi'_x\right) = -\frac{2r}{\left(r^2-1\right)s\sin 2\beta} = \frac{2r}{\left(1-r^2\right)s\sin 2\beta}\Bigg|_{-\beta}. \tag{13}$$

$$\Phi' = \left(\Phi'_y - \Phi'_x\right) = arctg\frac{2r}{\left(1-r^2\right)s\sin 2\beta}\Bigg|_{-\beta}. \tag{14}$$

Если же поворот системы координат осуществить на угол $+\beta$ как в [2], исходя из того же положения эллипса, описываемого каноническим уравнением (3), то система преобразования координат будет

$$\begin{aligned} x' &= x cos\beta + y\, sin\beta \\ y' &= -x sin\beta + y cos\beta. \end{aligned} \tag{15}$$

Или с учетом (3):

$$\begin{aligned} x' &= a\sqrt{cos^2\beta + r^2 sin^2\beta}\cdot\left(\frac{r sin\beta}{\sqrt{cos^2\beta + r^2 sin^2\beta}}cos\omega t - \frac{-cos\beta}{\sqrt{cos^2\beta + r^2 sin^2\beta}}sin\omega t\right), \\ y' &= a\sqrt{sin^2\beta + r^2 cos^2\beta}\cdot\left(\frac{r cos\beta}{\sqrt{sin^2\beta + r^2 cos^2\beta}}cos\omega t - \frac{sin\beta}{\sqrt{sin^2\beta + r^2 cos^2\beta}}sin\omega t\right). \end{aligned} \tag{16}$$

Введем обозначения: $$cos\Phi'_x = \frac{r sin\beta}{\sqrt{cos^2\beta + r^2 sin^2\beta}}, \quad sin\Phi'_x = \frac{-cos\beta}{\sqrt{cos^2\beta + r^2 sin^2\beta}}. \tag{17}$$

$$cos\Phi'_y = \frac{r cos\beta}{\sqrt{sin^2\beta + r^2 cos^2\beta}}, \quad sin\Phi'_y = \frac{sin\beta}{\sqrt{sin^2\beta + r^2 cos^2\beta}}. \tag{18}$$

Тогда $$tg\Phi'_y = \frac{1}{r}tg\beta, \quad tg\Phi'_x = -\frac{1}{r}ctg\beta. \tag{19}$$

$$tg\left(\Phi'_y - \Phi'_x\right) = \frac{2r}{\left(r^2-1\right)s\sin 2\beta}\Bigg|_{+\beta}. \tag{20}$$

$$\Phi' = \left(\Phi'_y - \Phi'_x\right) = arctg\frac{2r}{\left(r^2-1\right)s\sin 2\beta}\Bigg|_{+\beta}. \tag{21}$$

Связь старых и новых координат при повороте системы координат на угол $-\beta$ осуществляется по системе

$$\begin{aligned} x &= x' cos\beta + y' sin\beta \\ y &= -x' sin\beta + y' cos\beta. \end{aligned} \tag{22}$$

Подставим выражения для x и y в уравнение эллипса (2):

$$\frac{(x'cos\beta + y'sin\beta)^2}{E_x^2} - \frac{2(x'cos\beta + y'sin\beta)(-x'sin\beta + y'cos\beta)}{E_x E_y} cos\Phi +$$

$$+ \frac{(-x'sin\beta + y'cos\beta)^2}{E_y^2} = sin^2\Phi,$$

$$\left(\frac{cos^2\beta}{E_x^2}x'^2 + \frac{sin2\beta}{E_x^2}x'y' + \frac{sin^2\beta}{E_x^2}y'^2\right) + \left(\frac{sin2\beta cos\Phi}{E_x E_y}x'^2 - \frac{2cos2\beta cos\Phi}{E_x E_y}x'y' - \right.$$

$$\left. - \frac{sin2\beta cos\Phi}{E_x E_y}y'^2\right) + \left(\frac{sin^2\beta}{E_y^2}x'^2 - \frac{sin2\beta}{E_y^2}x'y' + \frac{cos^2\beta}{E_y^2}y'^2\right) = sin^2\Phi. \quad (23)$$

Произведем группировку слагаемых уравнения (23) относительно новых координат $x'^2, x'y', y'^2$.

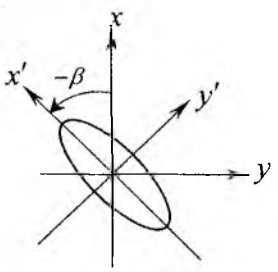


Рис.2

$$\left(\frac{cos^2\beta}{E_x^2} + \frac{sin2\beta cos\Phi}{E_x E_y} + \frac{sin^2\beta}{E_y^2}\right)x'^2 + \left(\frac{sin2\beta}{E_x^2} - \frac{2cos2\beta cos\Phi}{E_x E_y} - \right.$$

$$\left. - \frac{sin2\beta}{E_y^2}\right)x'y' + \left(\frac{sin^2\beta}{E_x^2} - \frac{sin2\beta cos\Phi}{E_x E_y} + \frac{cos^2\beta}{E_y^2}\right)y'^2 = sin^2\Phi. \quad (24)$$

Выбираем такой поворот осей координат на угол $-\beta$, который приводит уравнение эллипса к каноническому виду (рис.2). В таком случае следует приравнять нулю коэффициент при $x'y'$, т.е.

$$\frac{sin2\beta}{E_x^2} - \frac{2cos2\beta cos\Phi}{E_x E_y} - \frac{sin2\beta}{E_y^2} = 0. \quad (25)$$

или

$$sin2\beta\left(\frac{1}{E_x^2} - \frac{1}{E_y^2}\right) = \frac{2cos2\beta cos\Phi}{E_x E_y}, \quad sin2\beta\left(\frac{E_y^2 - E_x^2}{E_x^2 E_y^2}\right) = \frac{2cos2\beta cos\Phi}{E_x E_y}.$$

$$tg2\beta = \frac{2E_x E_y cos\Phi}{(E_y^2 - E_x^2)}. \quad (26)$$

Откуда находим

$$\beta = \frac{1}{2}arctg\frac{2E_x E_y cos\Phi}{(E_y^2 - E_x^2)}. \quad (27)$$

Запишем каноническое уравнение эллипса в новой системе координат:

$$\left(\frac{cos^2\beta}{E_x^2} + \frac{sin2\beta cos\Phi}{E_x E_y} + \frac{sin^2\beta}{E_y^2}\right)x'^2 + \left(\frac{sin^2\beta}{E_x^2} - \frac{sin2\beta cos\Phi}{E_x E_y} + \frac{cos^2\beta}{E_y^2}\right)y'^2 = sin^2\Phi. \quad (28)$$

Длины меньшей и большей полуосей эллипса определяются из его канонического уравнения приравниванием нулю соответственно x' и y', т. е. коэффициент эллиптичности

$$r=\frac{\dfrac{sin\Phi}{\sqrt{\dfrac{sin^2\beta}{E_x^2}-\dfrac{sin2\beta cos\Phi}{E_xE_y}+\dfrac{cos^2\beta}{E_y^2}}}}{\dfrac{sin\Phi}{\sqrt{\dfrac{cos^2\beta}{E_x^2}+\dfrac{sin2\beta cos\Phi}{E_xE_y}+\dfrac{sin^2\beta}{E_x^2}}}}=\sqrt{\frac{E_x^2sin^2\beta+E_xE_ysin2\beta cos\Phi+E_y^2cos^2\beta}{E_x^2cos^2\beta-E_xE_ysin2\beta cos\Phi+E_y^2sin^2\beta}}. \quad (29)$$

Если при данной ориентации поляризационного эллипса электромагнитной волны (рис. 2) осуществить поворот системы координат на угол $+\alpha$, равный $90^0-\beta$, то формулы (27) и (29) не изменятся.

Проделав аналогичные вычисления при повороте системы координат на угол $\alpha=-90^0+\beta$ и ориентации поляризационного эллипса, показанной на (рис. 3), получим следующие выражения для β и r:

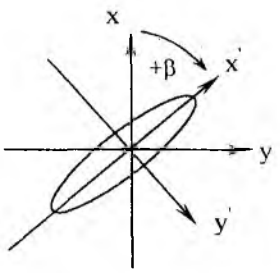


Рис.3

$$\beta=\frac{1}{2}arctg\frac{2E_xE_y\cos\Phi}{E_x^2-E_y^2}. \quad (30)$$

$$r=\sqrt{\frac{E_x^2\sin^2\beta-E_xE_y\sin2\beta\cos\Phi+E_y^2\cos^2\beta}{E_x^2\cos^2\beta+E_xE_y\sin2\beta\cos\Phi+E_y^2\sin^2\beta}}. \quad (31)$$

А при повороте системы координат на угол $+\beta$ выражения для β и r (30) и (31) не изменяются:

Возможен и другой подход к решению задачи анализа и синтеза поляризационного эллипса.

Возведем в квадрат оба уравнения (8) и составим их сумму и разность

$$E_x'^2=a^2\left(cos^2\beta+r^2sin^2\beta\right),$$

$$E_y'^2=a^2\left(sin^2\beta+r^2cos^2\beta\right).$$

$$E_x'^2+E_y'^2=a^2\left(1+r^2\right) \text{ или } r^2=\frac{E_x'^2+E_y'^2-a^2}{a^2}. \quad (32)$$

$$E_x'^2-E_y'^2=a^2\left[cos^2\beta+r^2sin^2\beta-sin^2\beta-r^2cos^2\beta\right]=$$

$$=a^2\left[cos2\beta-r^2cos2\beta\right]=a^2cos2\beta\left(1-r^2\right),$$

$$\left(1-r^2\right)=\frac{E_x'^2-E_y'^2}{a^2cos2\beta}. \quad (33)$$

Учитывая (32), получим:

$$E_y'^2=a^2\left(sin^2\beta+r^2cos^2\beta\right)=a^2sin^2\beta+a^2cos^2\beta r^2=$$

$$=a^2\sin^2\beta+\cancel{a^2}cos^2\beta\frac{E_x'^2+E_y'^2-a^2}{\cancel{a^2}},$$

$$E_y'^2=a^2sin^2\beta+\left(E_x'^2+E_y'^2-a^2\right)cos^2\beta.$$

Преобразуем это уравнение:

$$E'^2_y = a^2 \sin^2\beta + E'^2_x cos^2\beta + E'^2_y cos^2\beta - a^2 cos^2\beta,$$

$$E'^2_y - E'^2_y cos^2\beta = a^2 sin^2\beta + \left(E'^2_x - a^2 cos^2\beta\right),$$

$$E'^2_y sin^2\beta = sin^2\beta\left\{a^2 + \left(E'^2_x - a^2\right) ctg^2\beta\right\},$$

$$E'^2_y - a^2 = \left(E'^2_x - a^2\right) ctg^2\beta,$$

$$ctg^2\beta = \frac{E'^2_y - a^2}{E'^2_x - a^2},$$

$$\beta = arcctg\sqrt{\frac{E'^2_y - a^2}{E'^2_x - a^2}}. \qquad (34)$$

Ограничения, накладываемые на большую полуось эллипса a, следующие: 1) $a \neq 0$; 2) $a \le E'^2_y \le E'^2_x$.

Согласно (13)

$$tg^2\left(\Phi'_y - \Phi'_x\right) = \frac{4r^2}{\left(1-r^2\right)^2 \sin^2 2\beta}, \qquad (35)$$

Распишем знаменатель уравнения (35), учитывая (33)

$$\left(1-r^2\right)^2 \sin^2 2\beta = \frac{\left(E'^2_x - E'^2_y\right)^2}{a^4 \cos^2 2\beta} \sin^2 2\beta = \frac{\left(E'^2_x - E'^2_y\right)^2}{a^4} tg^2 2\beta. \qquad (36)$$

Известно, что

$$tg^2 2\beta = \frac{1}{\cos^2 2\beta} - 1. \qquad (37)$$

Преобразуем $\cos^2 2\beta$, учитывая (33) и (32)

$$\cos^2 2\beta = \frac{\left(E'^2_x - E'^2_y\right)^2}{a^4\left(1-r^2\right)^2} = \frac{\left(E'^2_x - E'^2_y\right)^2}{a^4\left(1 - \frac{E'^2_x + E'^2_y - a^2}{a^2}\right)^2} = \frac{\left(E'^2_x - E'^2_y\right)^2}{a^2\left[2a^2 - E'^2_x - E'^2_y\right]}. \qquad (38)$$

Тогда уравнение (36) преобразуется следующим образом:

$$\left(1-r^2\right)^2 \sin^2 2\beta = \frac{\left(E'^2_x - E'^2_y\right)^2}{a^4 \cos^2 2\beta} \sin^2 2\beta = \frac{\left(E'^2_x - E'^2_y\right)^2}{a^4} tg^2 2\beta.$$

$$\left(1-r^2\right)^2\sin^2 2\beta=\frac{\left(E'^2_x-E'^2_y\right)^2}{a^4}\left[\frac{a^2\left[2a^2-E'^2_x-E'^2_y\right]^2}{\left(E'^2_x-E'^2_y\right)^2}-1\right]=$$

$$=\frac{\left[a^2\left(2a^2-E'^2_x-E'^2_y\right)^2-\left(E'^2_x-E'^2_y\right)^2\right]}{a^4}. \qquad (39)$$

Тогда уравнение (35) примет вид

$$\mathrm{tg}^2\left(\Phi'_y-\Phi'_x\right)=\frac{4a^2\left(E'^2_x+E'^2_y-a^2\right)}{\left[a^2\left(2a^2-E'^2_x-E'^2_y\right)^2-\left(E'^2_x-E'^2_y\right)^2\right]}; \qquad (40)$$

С другой стороны,

$$\mathrm{tg}^2\left(\Phi'_y-\Phi'_x\right)=\frac{\sin^2\left(\Phi'_y-\Phi'_x\right)}{\cos^2\left(\Phi'_y-\Phi'_x\right)}=$$

$$=\sin^2\left(\Phi'_y-\Phi'_x\right)\left[\mathrm{tg}^2\left(\Phi'_y-\Phi'_x\right)+1\right],$$

$$\sin^2\left(\Phi'_y-\Phi'_x\right)=\frac{\mathrm{tg}^2\left(\Phi'_y-\Phi'_x\right)}{1+\mathrm{tg}^2\left(\Phi'_y-\Phi'_x\right)}. \qquad (41)$$

С учетом (40) уравнение (41) примет следующий вид:

$$\sin^2\left(\Phi'_y-\Phi'_x\right)=\frac{4a^2\left(E'^2_x+E'^2_y-a^2\right)}{\sqrt{\left[a^2\left(2a^2-E'^2_x-E'^2_y\right)^2-\left(E'^2_x-E'^2_y\right)^2\right]+4a^2\left(E'^2_x+E'^2_y-a^2\right)}}. \qquad (42)$$

Выводы

Таким образом, установлено, что в пределах одной определенной ориентации эллипса поляризации его параметры не зависят от направления поворота системы отсчета. Однако при левой ориентации (рис.2, когда β отрицательно) для определения β необходимо пользоваться формулой (27). Так как всегда $E_x \geq E_y$, то β в этом случае будет отрицательным, что соответствует действительности (рис. 2) и подстановка его в формулу (29) для определения коэффициента эллиптичности переводит ее в формулу (31). В этом случае можно говорить, что β и r инвариантны относительно направления поворота системы отсчета, но не относительно исходной ориентации эллипса поляризации.

Кроме того, в работе [2] допущена неточность при определении разности фаз между компонентами поля в новой системе координат при повороте ее на угол $+\beta$.

В предлагаемой статье получены общие выражения для определения параметров поляризационного эллипса, частный случай которых был исследован в работе [1].

Список литературы: 1. *Антенны* эллиптической поляризации. Теория и практика ; под ред.А.И. Шпунтова. – М. : Изд-во иностр. лит., 1961. – 335с. 2. *Канарейкин, Д.Б., Павлов, Н.Ф., Потехин,В.А.* Поляризация радиолокационных сигналов. – М. : Сов. радио, 1966. – 429с. 3. *Фрадин, А.З., Рыжков, Е.В.* Измерение параметров антенн. – М. : Связьиздат, 1962. – 316с.

Харьковский национальный университет им. В.Н. Каразина

Поступила в редколлегию 15.09.2012

УДК 004.056

Ю. В. ЛЫКОВ, канд. техн. наук

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТАЛЬНОГО ИССЛЕДОВАНИЯ ТЕЛЕФОННЫХ РАДИОЗАКЛАДНЫХ УСТРОЙСТВ МЕТОДОМ НЕЛИНЕЙНОЙ ЛОКАЦИИ

Введение

Среди всего многообразия способов несанкционированного перехвата информации особое место занимает прослушивание телефонных переговоров, поскольку телефонная линия – самый удобный и при этом самый незащищенный источник связи между абонентами в реальном масштабе времени. На начальном этапе развития телефонной связи никто особо не задумывался о защите линий от прослушивания, и электрические сигналы распространялись по проводам в открытом виде. В наше время прослушивание телефонной линии можно осуществлять, не заходя в помещение, при минимальных затратах и минимальном риске. Нужно просто подключить к телефонной линии (ТЛ) объекта специальное приемопередающее или регистрирующее устройство.

С точки зрения безопасности телефонная связь имеет еще один недостаток: возможность перехвата речевой (акустической) информации из помещений, по которым проходит ТЛ и где подключен телефонный аппарат. Это осуществимо даже тогда, когда не ведутся телефонные переговоры (так называемый микрофонный эффект телефона и метод высокочастотного навязывания). Для такого перехвата существует специальное оборудование, которое подключается к ТЛ внутри контролируемого помещения или даже за его пределами.

На данный момент разработано большое количество методов и средств защиты ТЛ, которые при их комплексном использовании позволяют достаточно эффективно закрыть данный канал утечки информации. Однако зачастую более целесообразно не защищать линию связи, а выявить сам факт несанкционированного доступа к ней, с последующим разоблачением злоумышленника, что позволит полностью исключить применение им других более совершенных методов (средств) съема информации. Для этих случаев разработаны различные средства контроля ТЛ, среди которых самыми эффективными являются рефлектометры и нелинейные локаторы.

Преимущества нелинейного локатора следуют из его способности обнаруживать:

- неработающие в данный момент закладные устройства (ЗУ);
- ЗУ с дистанционным управлением, находящиеся в режиме ожидания;
- ЗУ со специальными технологиями передачи информации, служащими повышению скрытности их работы (узкополосная модуляция, передача сигналов короткими сериями после их предварительного накопления в запоминающем устройстве, использование нескольких несущих частот, различные сложные виды модуляции и др.).

Постановка задачи

Известные публикации дают достаточно подробное изложение метода нелинейной локации (НЛ) при облучении электромагнитной волной полупроводниковых элементов, описание принципов работы устройств, реализующих этот метод, при этом практически отсутствует информация о средствах реализующих данный метод для поиска закладных устройств (ЗУ) именно в телефонных (проводных) линиях связи, которые имеют свою специфику применения.

Цель работы – рассмотрение метода нелинейной локации и экспериментальное исследование его возможностей для поиска радиозакладных устройств в телефонных линиях (ТЛ). Данная статья является продолжением работы [1], в которой изложены результаты моделирования метода НЛ в САПР OrCad 9.2.

Экспериментальное исследование нелинейных свойств ЗУ в ТЛ методом НЛ

В основе экспериментальной установки использовался персональный компьютер с ПО *Spectra Plus*. В качестве генератора гармонического сигнала применялась внешняя звуковая карта *Creative Sound Blaster Live 5.1 USB*. Сигнал отклика, снимаемый с резистора номиналом 51 *Ом*, подключенного последовательно в ТЛ, подавался на линейный вход звуковой карты. Линия передач состоит из двух участков (15 и 20 *м*) провода марки ТРП (диаметр токопроводящей жилы 0.5 *мм*) разъединённых ЗУ. В качестве ЗУ применялись различные физические модели (диоды, диодный мост, микросхемы) и малогабаритное подслушивающее устройство с питанием от ТЛ, смонтированное в одной из розеток для подключения телефонного аппарата. Второй конец ТЛ (ближний к АТС) при параллельном подключения к ней ЗУ размыкался, при последовательном – закорачивался. Функциональная схема экспериментальной установки показана на рис. 1

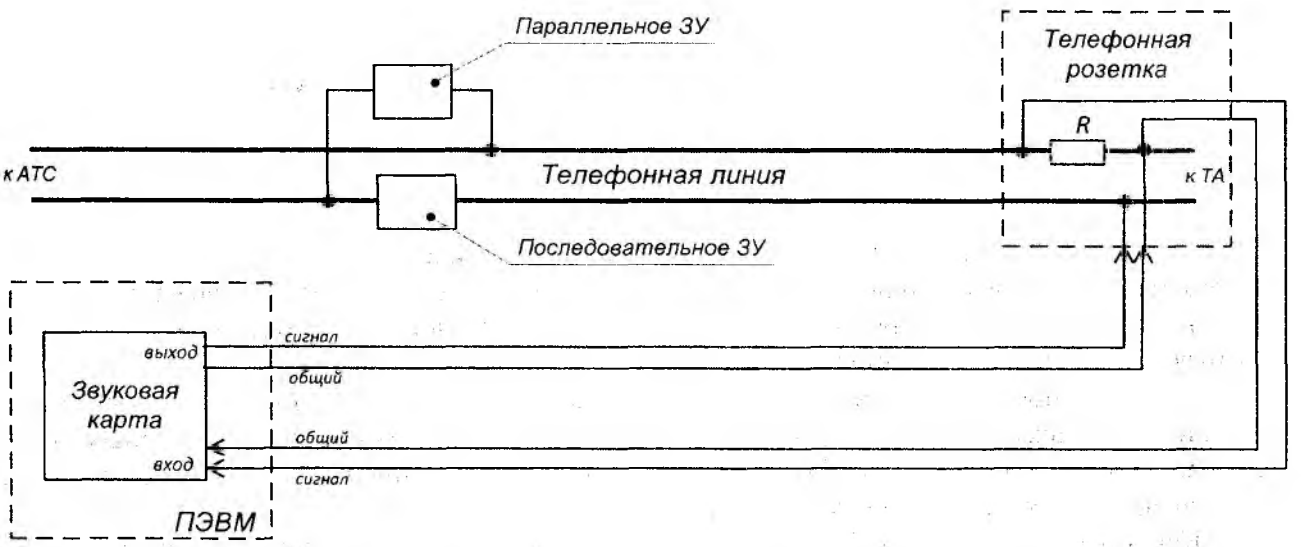


Рис. 1. Функциональная схема экспериментальной установки

До начала поиска ЗУ проведено исследование уровня шума в линии при отключенной АТС. Результаты показаны на рис. 2.

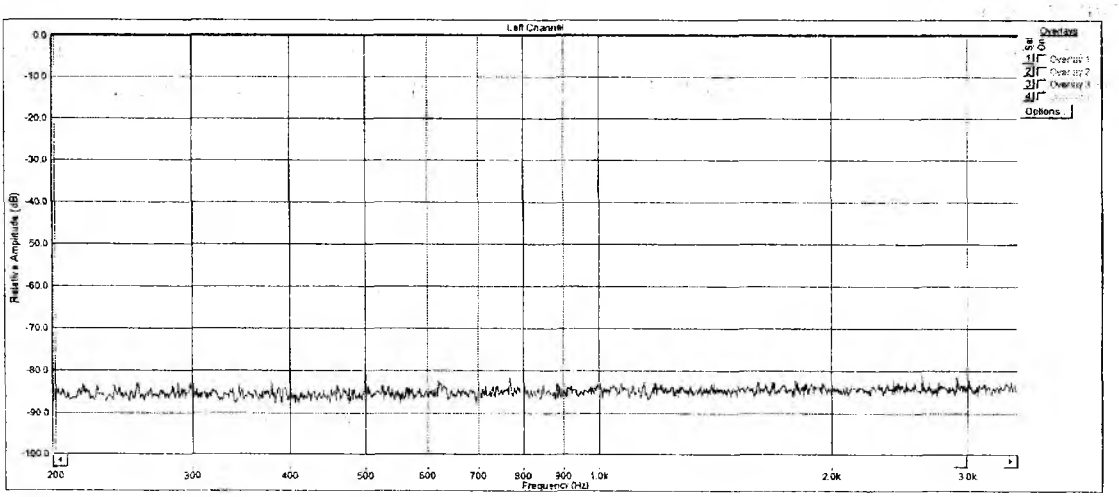


Рис. 2. Спектр шума в ТЛ

Как видно из рис. 2, уровень шума в исследуемой ТЛ не превышает 80 *дБ*.

Получен спектр отклика сигнала от НЛ в ТЛ подключенной к АТС (микроАТС К16010) результат показан на рис. 3.

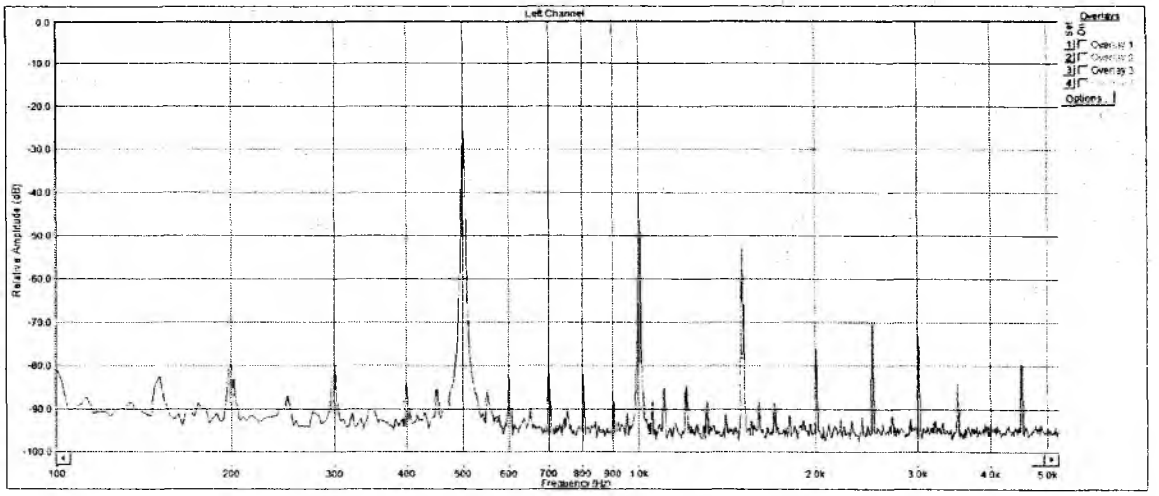


Рис. 3. Спектр сигнала в ТЛ при подключенной АТС

Проведенный эксперимент (рис. 3) показывает, что исследуемая на наличие закладны устройств линия обязательно должна быть отключена от АТС, которая вызывает значитель ные нелинейные искажения. Это подтверждает необходимость отключения всех известны нагрузок и источников питания от линии.

В работе проанализированы нелинейные свойства некоторых диодов как простейши моделей ЗУ, спектр отклика от диода Д2 показан на рис. 4. Для оценки нелинейных свойст выбран параметр *THD (Total Harmonic Distortion)* – коэффициент гармоник. Данный пара метр вычисляется автоматически программой *Spectra Plus*.

Как правило, на входе ЗУ находится диодный мост, упрощающий злоумышленнику подключение к ТЛ, поэтому было уделено внимание исследованию диодного моста как воз можной модели ЗУ. Кроме того, исследованы нелинейные свойства микросхем, которы могут применяться в ЗУ, в частности стабилизаторы напряжения 7805с и LM7905C.

Наиболее показательным является поиск реального ЗУ, которое последовательно под ключено к телефонной линии в месте соединения ТЛ с контактной розеткой и спрятан в ней. Данное ЗУ питается от ТЛ, его рабочая частота составляет 433 *МГц*, в схеме использу ется фильтр на ПАВ. Спектр отклика от данного закладного устройства показан на рис. 5.

Рис. 4. Спектр отклика от диода Д2

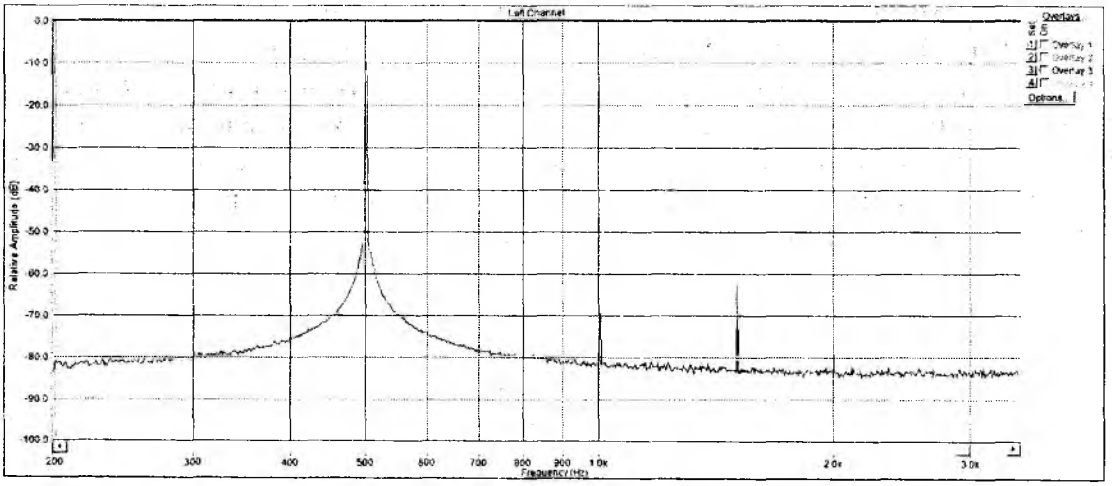


Рис. 5. Спектр отклика от ЗУ

Как видно из рис. 5, первая гармоника на 60 *дБ* превышает вторую и на 50 *дБ* – третью, коэффициент нелинейных искажений равен 0,22 %. Несмотря на незначительное значение коэффициента гармоник, по полученной спектрограмме можно достаточно точно идентифицировать наличие нелинейного элемента. Значения параметра нелинейных искажений для каждого исследуемого объекта сведены в таблицу:

Объект исследования	Коэффициент гармоник
Диод Д2	1,41
Диод Д9	1,52
Диод Д226	1,60
Диодный мост КЦ406	0,73
Микросхема 7805с	1,28
Микросхема LM7905C	0,54
Телефонное закладное устройство №1	0,01
Телефонное закладное устройство №2	0,32

В результате экспериментального исследования, а также моделирования [1] были обнаружены все используемые в работе телефонные ЗУ и их простые физические и математические модели.

Возможности метода НЛ наглядно продемонстрированы на рис. 6. Максимально неуязвимыми для устройств контроля ТЛ, в том числе и для НЛ, являются ЗУ с индуктивным подключением, которое в свою очередь может быть реализовано через трансформатор или магнитную рамку. При использовании достаточно высоких напряжений (для кабеля ТРП,ТРВ, испытательное напряжение в течение 1 *мин* составляет 1000 *В* [3]) зондирующего сигнала возникает потенциальная возможность и по обнаружению ЗУ с индуктивным подключением к ТЛ.

Кроме того, методом НЛ возможно обнаруживать устройства съема акустической информации в помещении, реализующие метод ВЧ-навязывания.

Разработка требований к нелинейному локатору

Основными параметрами нелинейного локатора являются: частота зондирующего (выходного) сигнала; выходное напряжение; коэффициент гармоник выходного сигнала; динамический диапазон измеряемых уровней сигнала на частотах гармоник.

В качестве частоты зондирующего сигнала нелинейного локатора целесообразно выбирать частоты менее 20 *кГц*, так как при этом обеспечивается максимальное отношение сигнал/шум. Кроме того, для минимизации влияния помех от сети питания 220 *В* необходимо чтобы частоты гармоник были не кратны частоте 50 *Гц*. Для повышения достоверности получаемых результатов необходимо предусмотреть возможность перестройки нелинейного локатора по частоте. С другой стороны, для повышения скрытности проведения поисковых работ частота зондирующего сигнала должна быть выше 16 *кГц*.

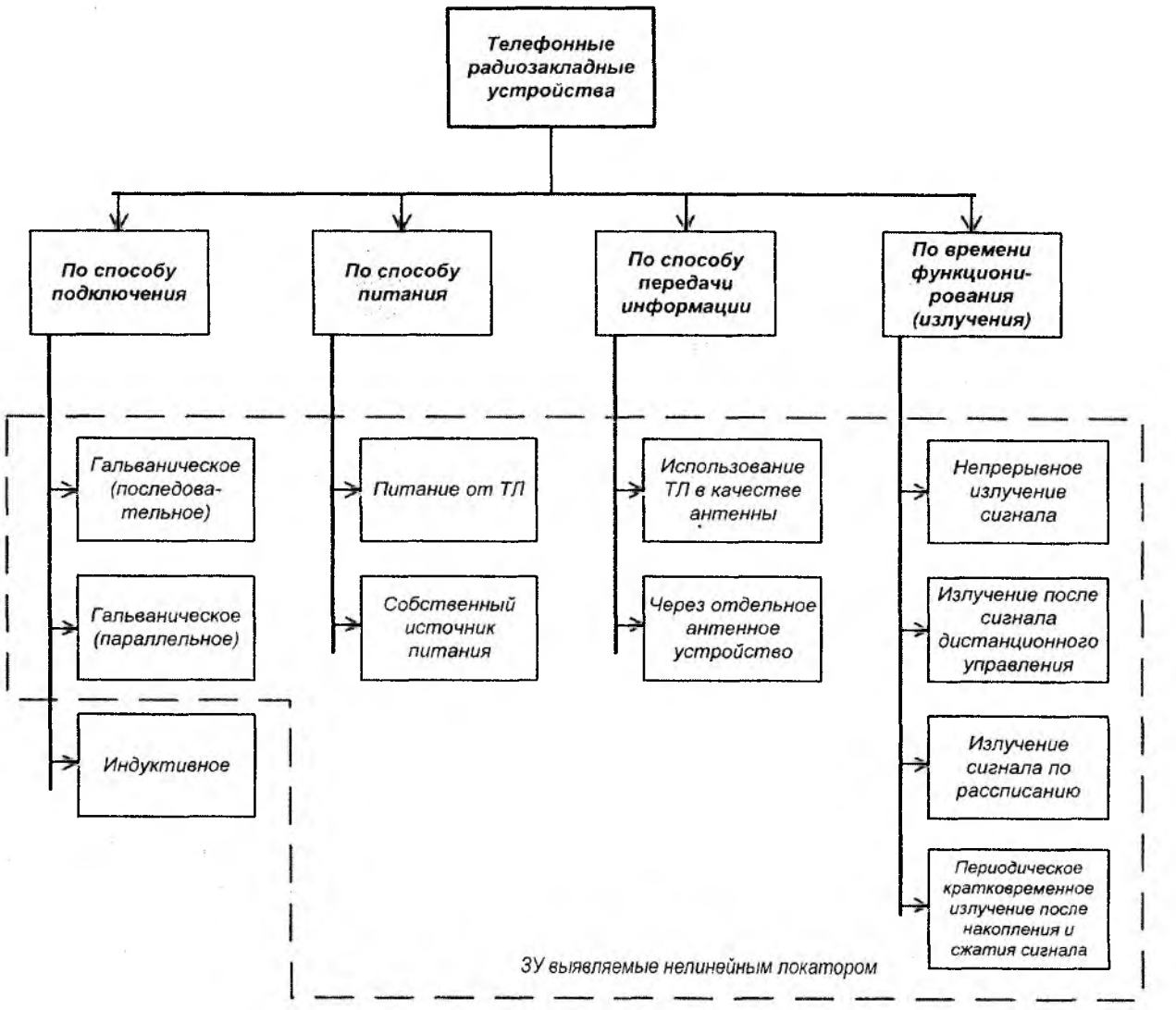


Рис. 6. Классификация телефонных радиозакладных устройств

Исходя из значений рабочего затухания абонентской линии телефонной сети на частоте 1000 *Гц*, которое согласно [4] должно быть не более 5,0 *дБ* и уровнем тепловых шумов в линии -90 *дБВ* (для линии длиной 500 *м*, полосы частот 20 *кГц*, комнатой температуры), а также возможного дополнительного затухания (1 – 5 *дБ*), вызванного подключенным ЗУ, для обеспечения динамического диапазона измерения 70 *дБ* (рис. 5) необходимо максимальное значение выходного напряжения нелинейного локатора порядка 1 – 10 *В*.

Коэффициент гармоник зондирующих сигналов согласно данным эксперимента и моделирования должен быть менее 0,1 %. Для его уменьшения, например при использовании некачественной внутренней звуковой карты ПЭВМ, можно использовать режекторные фильтры, настроенные на частоты второй и третьей гармоник.

Для активации некоторых ЗУ и введение их в рабочий режим в схеме должен быть предусмотрен источник постоянного напряжения на 10 – 15 *В*. В результате обобщённая функциональная схема нелинейного локатора показана на рис. 7.

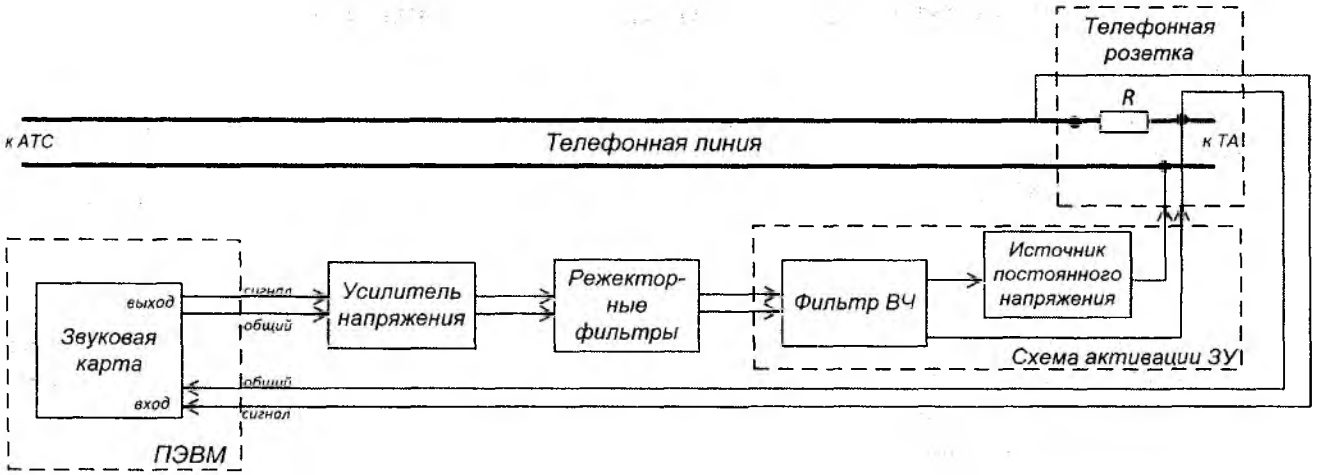


Рис. 7. Обобщённая функциональная схема нелинейного локатора

Выводы

В работе исследован метод нелинейной локации применительно к поиску радиозакладных устройств в проводных телефонных линиях связи. В результате натурных испытаний получено следующее:

- диапазон коэффициентов гармоник для физических моделей ЗУ и реальных ЗУ лежит в пределах от 0,01 до 1,5;

- при небольших протяженностях линии связи и малом уровне помех в качестве аппаратной реализации нелинейного локатора может использоваться ноутбук с качественной звуковой картой, что значительно доступнее, чем применение специализированных нелинейных локаторов. При исследовании протяженных ТЛ или/и высоком уровне помех может потребоваться внешний усилитель на выходное напряжение 10 *В*;

- применение метода нелинейной локации телефонной линии позволяет обнаружить подключение большинства используемых злоумышленниками типов телефонных закладных устройств.

Список литературы: 1. *Лыков, Ю.В.* Поиск радиозакладных устройств в телефонной линии методом нелинейной локации / *Ю.В. Лыков* // Восточно-Европейский журнал передовых технологий. – Харьков : Технологический центр, 2012. – №6/9 (60). – С.56-72. 2. *Вернигоров, Н.С.* Принцип обнаружения объектов нелинейным локатором // Защита информации. Конфидент. – 1998. – №4. –С. 65-70. 3. *ГОСТ* Р 51311-99. Кабели телефонные с полиэтиленовой изоляцией в пластмассовой оболочке. Технические условия. – Введ. 2000-07-01. – М. : Изд-во стандартов, 2000. – 27 с. 4. *ОСТ 45.83-96* Сеть телефонная. Линий абонентские кабельные с металлическими жилами. Нормы эксплуатационные. – Введ. 1997-22-12. – М. : Изд-во стандартов, 1997. – 15 с.

Харьковский национальный университет радиоэлектроники

Поступила в редколлегию 13.09.2012

УДК 551.501.7

В.И. ЛЕОНИДОВ, канд. техн. наук

МОДЕЛЬНО-СТРУКТУРНЫЙ АНАЛИЗ ЭХОСИГНАЛОВ АКУСТИЧЕСКОГО ЗОНДИРОВАНИЯ АТМОСФЕРЫ

Введение

Объектом наблюдения акустического локатора являются турбулентные моли в поле температуры и скорости, которые создают неоднородности волнового сопротивления воздушной среды. В общем случае цели такого типа относятся к объемно-распределенным, так как в импульсном объеме может наблюдаться более одного объекта рассеяния акустической волны.

Турбулентные моли формируются в процессе теплообмена между подстилающей поверхностью и атмосферой, и поэтому их «плотность», теплосодержание и температурный контраст определяются режимом баланса тепла между подстилающей поверхностью и атмосферой. Следовательно, характер цели акустического локатора зависит от типа термодинамического состояния атмосферы.

Относительно модели «мгновенной цели» акустического локатора выдвигаем следующую гипотезу: цель представляет собой горизонтально-слоистую среду, которая характеризуется «скачками» волнового сопротивления по вертикали, при этом горизонтальными градиентами в пределах импульсного объема можно пренебречь.

В структуре цели выделим переходные зоны (зона скачка) и зоны слоев. В пределах зоны слоя, ограниченного импульсным объемом, будем считать волновое сопротивление среды постоянной величиной $\rho_в = const$.

Вертикальная протяженность ∂h_i (толщина) зоны скачка отвечает условию $\partial h_i << \lambda_s$. Приращение температуры dT_i в переходном слое отвечает неравенству $dT_i >> \sigma_{Tв}$, а градиент температуры в зоне «скачка» $dT_i/dh_i >> \gamma_a$, где λ_s – длина акустической волны, $\sigma_{Tв}$ – ср.кв. отклонение поля температуры, γ_a – сухоадиабатический температурный градиент.

При таких допущениях параметры эхосигналов, полученных при реальном акустическом зондировании, представляют собой случайную последовательность импульсов, длительность которых не меньше длительности зондирующего импульса.

Цель настоящей работы состоит в создании и апробации методики модельно-структурного анализа эхосигналов, которая позволит изучать свойства турбулентных молей температурного поля атмосферы как целей акустического локатора. Работа является дальнейшим развитием исследований, начатых в этой области в работах [1 – 2]. Экспериментальный материал, пример которого анализируется в настоящей работе, был получен автором на станции акустического зондирования атмосферы, созданной в Харьковском национальном университете радиоэлектроники. Экспериментальные наблюдения проводились в летние периоды времени 2011 и 2012 годов. На рис.1 приведен внешний вид антенных систем станции акустического зондирования атмосферы.

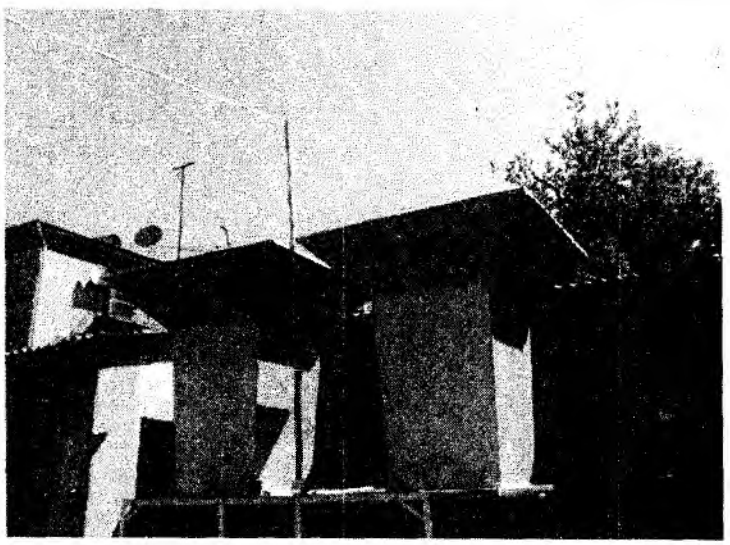


Рис.1 Антенные системы станции акустического зондирования атмосферы

Модельно-структурный анализ эхосигналов

Результатом акустического зондирования атмосферы является реализация эхосигналов, организованная в виде $(m \times n)$ матрицы S_a цифровых отсчетов эхосигналов $s_{i,j}$, где $i = 1 \ldots m$, $j = 1 \ldots n$, причем столбцы матрицы представляют собой вертикальные одиночные профили эхосигналов $s_j(i)$ и, следовательно, каждая строка $s_i(j)$ матрицы S_a представляет собой ряд дискретных отсчетов эхосигналов для данной высоты $h_i = \Delta h \cdot i$, где Δh – дискретный шаг высоты.

Так как $h = c_s \cdot t$, где c_s – скорость распространения акустических волн в атмосфере, t – время, то профиль эхосигналов есть также и функция времени $s_j = F(t)$.

На рис. 2 приведен фрагмент реализации эхосигналов, продолжительностью $T_\Delta \cong 10$ *мин*, что позволяет визуализировать детальную структуру эхосигнала.

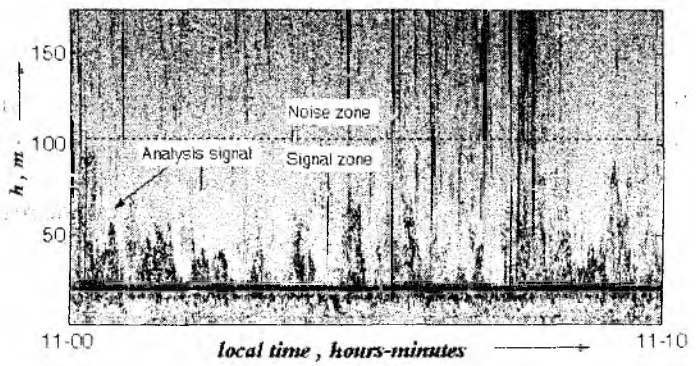


Рис.2. Фрагмент эхограммы акустического зондирования

Целью анализа является выявление структурных особенностей реализаций эхосигналов. Предположим, что реализация $s_j(i)$ эхосигналов есть случайная последовательность эхоимпульсов, принятая системой за один период повторения T_s зондирующих импульсов $T_s = \Delta t \cdot m$, где Δt – интервал времени на дискретный шаг высоты. Для принятой реализации $s_j(i)$ определим статистические частоты p^* возникновения импульсных эхосигналов s_τ, длительностью $\tau_e \geq \tau_r$, где τ_e – длительность зондирующего импульса, τ_r – длительность импульсных эхосигналов в принятой последовательности.

Расчет проведем методом модельно-корреляционного анализа реализации $s_j(i)$.

Примем модель одиночного импульса в принятой реализации в виде

$$s_p = -\cos\left(2\pi \cdot x_p / T_p\right) \qquad (1)$$

где $x_p = 1 \ldots p$, p – число дискретных отсчетов на длительность модели, T_p – период модельной функции в дискретных отсчетах, $p = k_p \cdot T_p$, где k_p принимает значения $k_p = 1; 2$; T_p – переменная величина, ее значения определяются из выражения $T_p(n_p) = k_p \cdot (T_* + n_p)$, где $n_p = 1 \ldots n_m$, n_m – максимальное число итераций изменения периода модельной функции, k_p – постоянный масштабный коэффициент.

Далее, для каждого из значений T_p рассчитываются коэффициенты корреляции $k_r(q)$ между моделью s_p и фрагментом реализации эхосигнала $s_j(i)$, длительность которого $T_{\Delta s}$ отвечает условию $T_{\Delta s} = p$:

$$k_r\left(q,\ T_p = const\right) = \frac{1}{\sigma_s \cdot \sigma_p} \sum_{i=q}^{q+p-1} s_j^{\circ}(i) \cdot s_p^{\circ}(i-q+1) \qquad (2)$$

где $q = 1 \ldots (m-p)$, $\sigma_{s,p}$ – ср.кв.отклонения фрагмента реализации эхосигналов и модели импульсного эхосигнала соответственно.

Каждый отсчет полученной зависимости $k_r(q)$ сравнивается с пороговым значением g и формируется $(n_m \times m)$ матрица B, члены строк которой принимают значения из условия $B(i, j) = 1$ при $k_r \geq g$ и $B(i, j) = 0$ при $k_r < g$.

Если принять $g = 0{,}8$, то можно считать, что сумма членов строк матрицы B представляет собой число случаев приема эхосигналов в виде импульсов, определяемых моделью s_p. Тогда статистическая частота определится из выражения

$$p^*(i) = \frac{1}{m - p_i} \sum_{j=1}^{n_m} B_i(j) \qquad (3)$$

Для анализа был выбран фрагмент эхограммы, указанный на рис.2 стрелкой. Был проведен структурный анализ реализаций эхосигналов и помех. Для этого соответствующие реализации формировались по областям эхограммы, указанным на рис.2 разделительной линией.

Результаты расчетов приведены на рис. 3, *а*, *б* и 4, *а*, *б* соответственно для двух видов моделей. В правом верхнем углу рисунков приведены графики функций, описывающих модели. На рис. 3, *а* и 4, *а* приведены распределения сигналов, на рис. 3, *б* и 4, *б* – помех. Ординаты графиков изображены в виде полос, напоминающих гистограмму для удобства восприятия величины p^*.

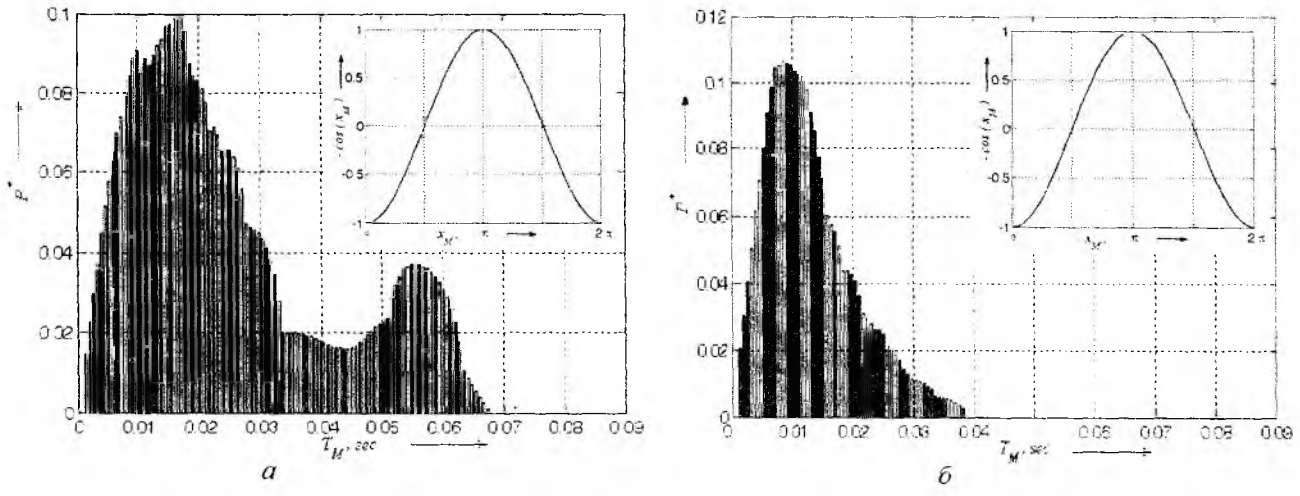


Рис.3. Рапределение статистической вероятности реализации эхосигналов

Выводы

Полученные результаты показывают, что при использовании модели в виде одного периода косинусного сигнала статистическая вероятность появления такого вида сигналов имеет вид сплошного распределения в области значений длительностей импульсов $\tau_r = (2 \div 68) \cdot 10^{-3}\, c$ (длительность зондирующего импульса $\tau_e = 2{,}9 \cdot 10^{-3}\, c$). Из этого следует, что в импульсном объеме наблюдаются «скачки» волнового сопротивления среды с интервалами $\delta h_i << \Delta h$, где Δh – разрешающая способность системы по дальности, для данной системы зондирования $\Delta h \cong 0{,}5\, м$.

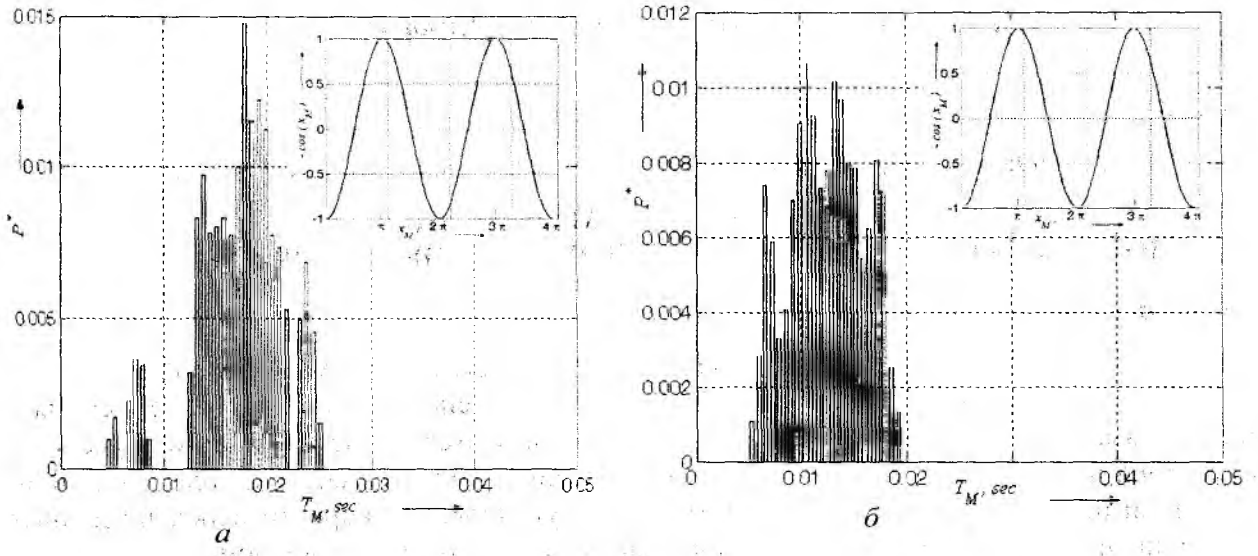


Рис.4. Распределение статистической вероятности реализации эхосигналов

Распределение помехи примерно вдвое уже, чем распределение сигналов. Это говорит о том, что в черте мегаполиса превалируют импульсные акустические помехи.

При использовании модели в виде двух периодов косинусного сигнала распределение сигналов становится линейчатым. Следовательно, упорядоченные структуры существуют в турбулентном поле температуры только при определенных масштабах длины.

Распределение помех в этом случае так же, как и в предыдущем случае, близко к сплошному, но в более узкой полосе периодов.

Основным результатом работы является создание методики структурного анализа эхосигналов акустического зондирования и формализованное подтверждение гипотезы о пространственно распределенном характере цели – турбулентных молей в поле температуры.

Выводы относительно структуры полей акустических помех и сигналов следует рассматривать как предварительные, так как они следуют из обработки относительно небольшого объема экспериментальных данных.

Из полученных результатов также следует, что перспективным направлением в развитии аппаратной части систем акустического зондирования является повышение их разрешающей способности по дальности, что позволит измерить мощность эхосигнала, рассеянного отдельным турбулентным молем.

Список литературы: 1.*Семенец, В.В.* Акустическое зондирование атмосферы в проблеме изучения процессов теплообмена в зоне мегаполиса / В.В. Семенец, В.И. Леонидов // Восточно-европейский журнал передовых технологий. – 2011. – 6/8(54). – С.54. 2. *Леонидов, В.И.* Анализ квазипериодических амплитудно-временных структур эхосигналов акустического зондирования атмосферы / В.И. Леонидов, В.В. Семенец // Радиотехника. – 2011. – Вып.167. – С. 27 -31.

Харьковский национальный
университет радиоэлектроники

Поступила в редколлегию 07.09.2012

СИСТЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

УДК 621.391

И.С. ШОСТКО, д-р техн. наук, Ю.Э. СОСЕДКА, АЛМАКАДМА ТАХА

РАЗРАБОТКА РЕКОМЕНДАЦИЙ ПО РЕГУЛИРОВАНИЮ ПРОПУСКНОЙ СПОСОБНОСТИ В WPAN

Введение

Мир телекоммуникаций и передачи данных сталкивается с динамично растущим спро сом на частотные и временные ресурсы. В беспроводных системах связи, эта тенденция і основном связана с увеличением числа мобильных пользователей Internet и также с увеличе нием объемов передаваемой информации. Полоса пропускания в расчете на одного пользо вателя стремительно увеличивается. Поэтому задача увеличения скорости передачи данны (пропускной способности) в беспроводных каналах связи является актуальной.

Одним из перспективных направлений развития беспроводной связи для WPAN являет ся переход к сверхширокополосным (СШП) сигналам (технология UWB Multiband OFDM (англ. Orthogonal frequency-division multiplexing)) [1], которые обеспечивают максимальную на сегодняшний день скорость передачи данных в радиоканале, высокую помехоустойчи вость в условиях многолучевого распространения и воздействия помех (стандарт IEEE 802.15.3a) [2]. Для увеличения скорости передачи данных в радиоканале и обеспечения вы сокой помехоустойчивости в условиях многолучевого распространения и воздействия поме необходимо в перспективных системах WPAN использовать технологию UWB Multiband OFDM при оптимальном управлении параметрами сигнала.

Описание технологии Wireless Universal Serial Bus (WUSB)

Технология стандартизирована институтом IEEE – стандарт IEEE 802.15.3a [3]. Рабочие частоты – 3,1 – 10,6 *ГГц*, поддерживаемая скорость передачи данных: 53.3, 80, 106.7, 160 200, 320, 400, 480, 640, 800, 960 и 1024 *Мбит* в секунду.

Однако расстояние от передатчика до приемника, при котором обеспечивается надеж ная передача данных со столь высокой скоростью, весьма мало. В пределах трех метров га рантируется скорость 480 *Мбит/с*. При увеличении этого расстояния до 10 *м* максимальна скорость передачи информации падает до 110 *Мбит/с*, и при дальнейшем возрастании рас стояния от передатчика до приемника эта скорость продолжает стремительно убывать (при чем гораздо быстрее, чем при использовании иных технологий беспроводной передачи дан ных — Wi-Fi, Bluetooth и др.)

Транспортный уровень Wireless USB базируется на UWB технологии; физический пред ставляет собой уровень формирования среды передачи данных, где помимо WUSB с легко стью могут фигурировать W1394 (Wireless FireWire), Bluetooth и прочие.

Особенности протокола Wireless USB

Структура кадра [2], который состоит из трех компонентов: преамбула PLCP, заголово PLCP и PSDU. Компоненты перечислены в порядке передачи. Преамбула PLCP содержи пакет/синхронизации кадровой последовательности и канала оценки последовательности PSDU является последним компонентом кадра. Этот компонент формируется из полезно нагрузки, кадра проверки (FCS), 6 *бит* и, бит, которые вставлены с целью обозначения гра ницы чередования символов. Полезная нагрузка содержит фактические данных для переда чи. Полезная нагрузка может варьироваться от 1 до 4095 *байт* или даже может быть полно стью опущена, если она не является полезной в данном контексте. Когда необходимо пере дать больше, чем 4095 *байт*, общую сумму делят на несколько кадров. В конце полезно нагрузки добавляется проверочный код FCS, чтобы помочь обнаружить ошибки передачи.

Участок из 6 *бит* влияет только на скорость передачи данных. При передаче пакетов преамбула PLCP отправляется первой, затем PLCP заголовок и, наконец, PSDU. Систематическая часть заголовка PLCP всегда передаётся со скоростью 39,4 *Мбит/с*. PSDU передаётся с желаемой скоростью передачи данных 53,3, 80, 106,7, 160, 200, 320, 400, 480, 640, 800, 960 или 1024 *Мбит/с*.

Проведём оценку накладных расходов, связанных с передачей данных, и вычислим эффективную пропускную способность. Накладные расходы для 480 *Мбит/с* приведены на рис. 1.

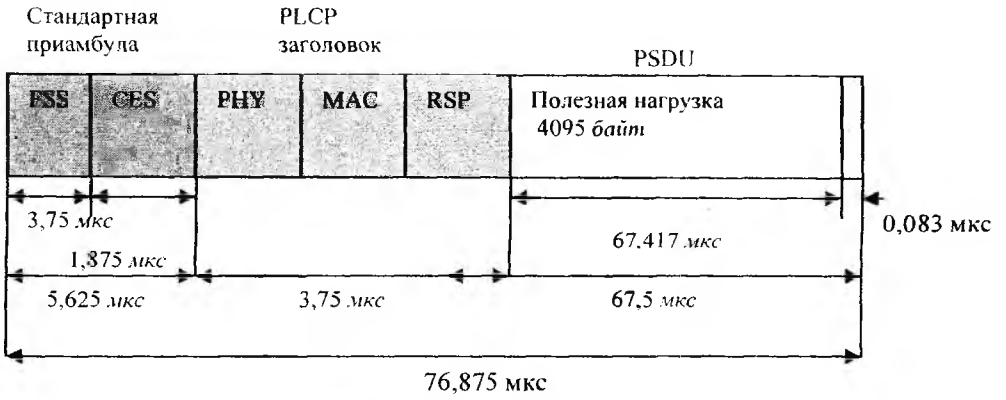


Рис.1. Структура временных интервалов кадра для 480 *Мбит/с*

Оценка накладных расходов времени:

$$\frac{t_{общ} - t_{полезной\ нагрузки}}{t_{общ}} = \frac{76{,}875 - 67{,}417}{76{,}875} = 12{,}3\% \qquad (1)$$

или эффективный эквивалентный битрейт 420,9 *Мбит/с*.

То же самое можно вычислить для 53,3 *Мбит/с*, как показано на рис. 2.

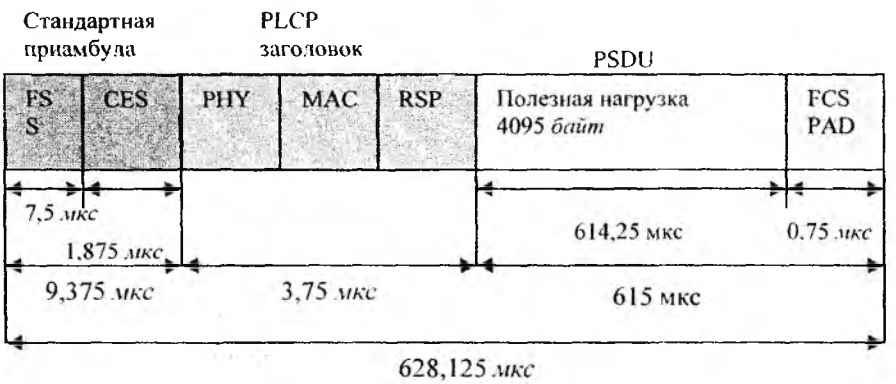


Рис. 2. Структура временных интервалов кадра для 53,3 *Мбит/с*

Оценка накладных расходов времени:

$$\frac{t_{общ} - t_{полезной\ нагрузки}}{t_{общ}} = \frac{628{,}125 - 615}{628{,}125} = 2{,}09\% \qquad (2)$$

или эффективный эквивалентный битрейт 52,2 *Мбит/с*.

Результаты расчёта эффективности выбранного протокола η_p для максимальной длины полезной нагрузки для каждого стандарта номинальной скорости передачи данных, для

стандартного кадра (1) и с разрывом преамбулы (2) (допускается только для скорости передачи данных выше, 200 *Мб/с*) показаны на рис. 4.

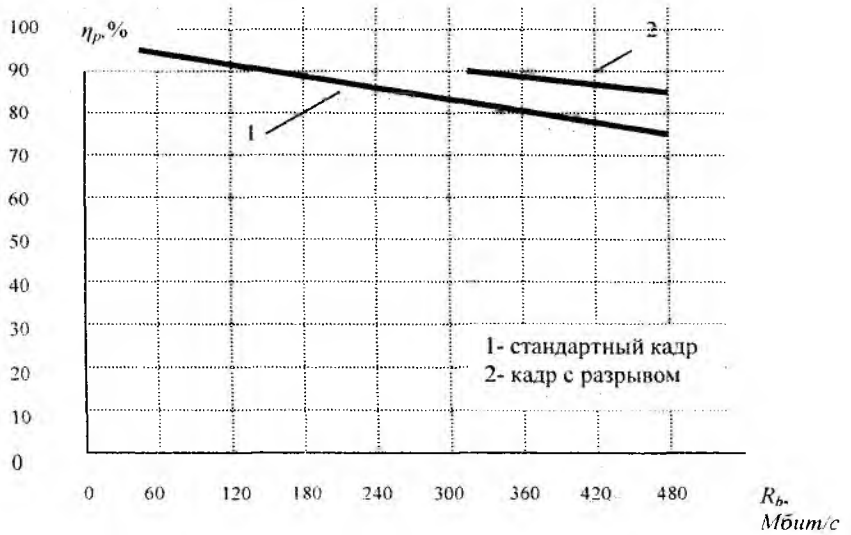


Рис. 3. Эффективность протокола η_p как функция от R_b

Разработка модели UWB Multiband OFDM приёмопередатчика

Имитационная модель физического уровня приёмо-передающей системы UWB была разработана в среде Matlab/Simulink [4 – 6]. Блок-схема данной модели приведена на рис. 4.

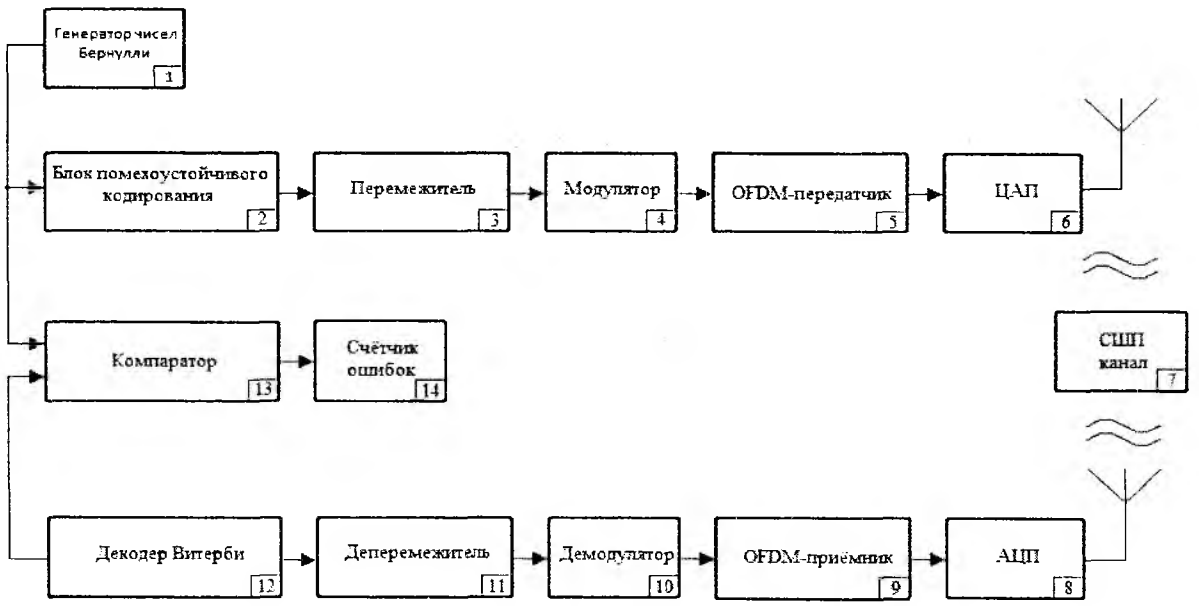


Рис. 4. Блок-схема модели СШП физического уровня

Первый блок представляет собой генератор чисел Бернулли, для получения случайной последовательности бит.

Второй блок – блок помехоустойчивого кодирования детальнее представлен на рис. 5:

Рис. 5. Блок-схема помехоустойчивого кодирования в системе UWB

Третий блок представляет собой перемежитель. Основная его цель состоит в том, чтобы сделать потенциальную групповую ошибку одиночной в нескольких местах, это даёт возможность увеличить возможность правильного декодирования сигнала.

Четвёртый блок представляет собой модулятор, в нашей работе модель собрана на основе QPSK модуляции. Этот вариант модуляции выбран на основе проведённых исследований [7], где на основе опытного сравнения доказано, что система MB-OFDM UWB эффективнее работает с QPSK модуляцией, в которой используется созвездие из четырёх точек, расположенных на равных расстояниях (рис. 6).

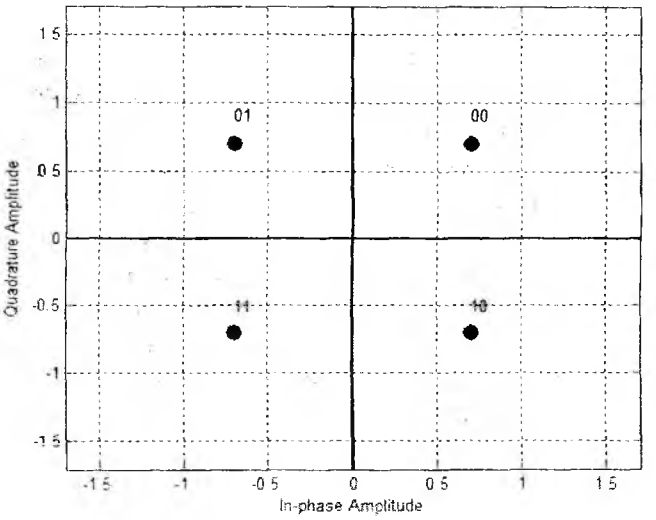


Рис. 6. Сигнальное созвездие QPSK

Для QPSK модуляции вероятность битовой ошибки имеет следующий вид

$$erfc(x) = 1 - erf(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt \tag{3}$$

$$BER_{qpsk} = \tfrac{1}{2} \cdot erfc^2 \left(\sqrt{\frac{E_s}{N_o}} \right) \tag{4}$$

где *erfc* – дополнительная функция ошибок, которая в свою очередь имеет вид

$$erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt \tag{5}$$

На рис. 7 отображен пятый блок модели приёмо-передающей системы UWB – OFDM передатчик:

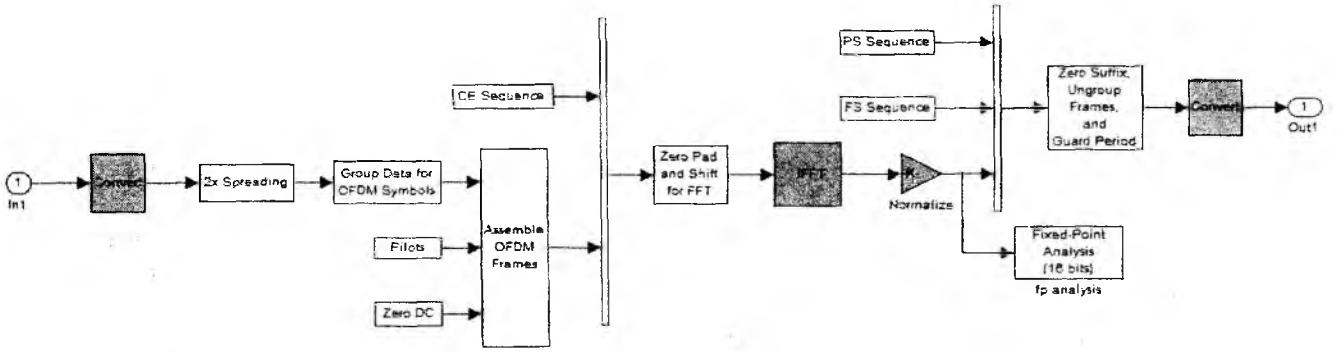


Рис. 7. Общая структура OFDM-передатчика

Цифровой сигнал, преобразованный в системе кодеров, поступает на полосовой модулятор. Выходная последовательность с модулятора поступает в мультиплексор, где и разбивается на множество потоков, которые будут поступать на блок обратного быстрого преобразования Фурье (ОБПФ). Далее необходимо реализовать мультиплексирование потока данных по частоте. Эту операцию выполняют следующим образом: последовательный поток преобразуют в матрицу, каждый столбец которой далее будет преобразован в OFDM сигнал. После ОБПФ нужно создать циклический префикс (ЦП). Но передавать такой сигнал нельзя, его необходимо усилить.

При использовании квадратурно-амплитудной модуляции с количеством точек на созвездии большим, чем у QAM-4(QPSK), эти точки не будут соответствовать постоянной мощности, вследствие чего появляется такое понятие, как PAPR (Peak to Average Power Ratio) или отношение пиковой мощности к средней мощности сигнала. Для уменьшения значения PAPR используется метод Резервирования Тона (Tone Reservation (TR)). Он предполагает выделение дополнительных поднесущих с информацией, которая минимизирует PAPR.

В TR основная идея состоит в том, чтобы зарезервировать некоторые поднесущие OFDM, называемые Пиковые Тоны Сокращения (PRT) для снижения PAPR. Эти поднесущие не несут информации о данных, и используются только для того, чтобы уменьшить PAPR. Этот метод ограничивает вектор данных, и сокращает пиковый вектор, при этом не имеет связи с сеткой частот. Из-за этого условия в сигнале не возникают нелинейные искажения. Значения, использованные для преобразования сигнала, просто кодируются и передаются. Это объясняет увеличение вероятности битовой ошибки при использовании модуляции порядка выше четырёх.

Проблема заключается в том, что если сигнальное созвездие представляет собой квадратную сетку, символьная ошибка происходит при выходе за пределы квадрата со стороной, равной шагу квадратной сетки. При BPSK (рис. 8, *а*) и QPSK (рис. 8, *б*) все значения являются краевыми, поэтому рост вероятности ошибки линеен. Но в варианте QAM-8 (рис. 8, *в*) появляется понятие «краевой эффект», когда для краевых точек вероятность ошибки меньше, чем для остальных. Отсюда появляется нелинейное возрастание вероятности ошибки при порядках модуляции выше четырёх.

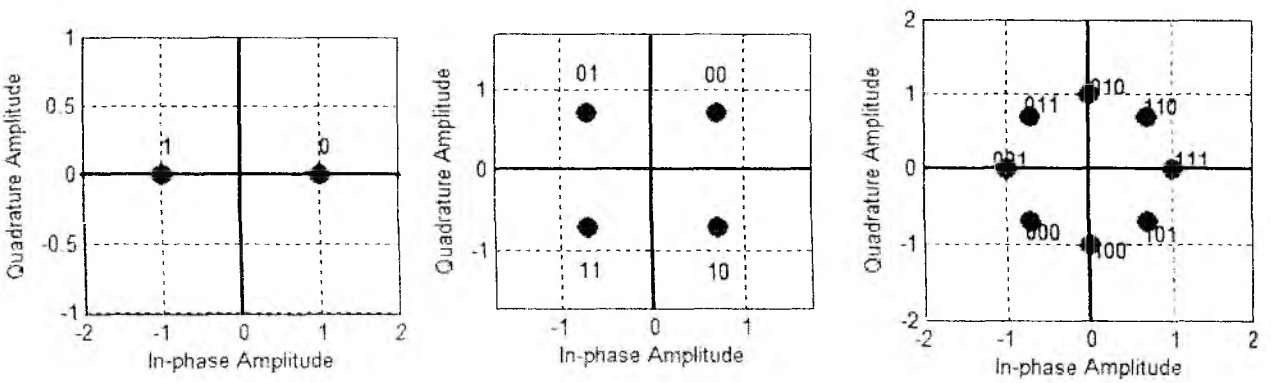


Рис.8. Сигнальные созвездия BPSK – *а*, QPSK – *б* и QAM-8 – *в*

С выхода OFDM-модулятора цифровой сигнал преобразуется в аналоговый посредством ЦАП (блок 6, рис. 4). После чего посредством СШП антенны, сигнал излучается в общую среду передачи (блок 7 – СШП канал, рис. 4). В приёмную часть входят: АЦП OFDM-приёмник, QAM-демодулятор, деперемежитель, декодер Витерби, компаратор и счётчик ошибок.

В конечном итоге, UWB модель позволяет проделать следующее:

– исследовать зависимость вероятности битовой ошибки от скорости свёрточного кодирования;

– исследовать зависимость вероятности битовой ошибки от особенностей канала связи (прямая видимость сигнала, приём с переотражениями – CM1-4);

– исследовать зависимость вероятности битовой ошибки от расстоянии между приёмником и передатчиком;

– исследовать зависимость вероятности битовой ошибки от размера символа OFDM.

Анализ результатов моделирования и разработка рекомендаций по увеличению скорости передачи данных в WPAN с использованием СШПС

Задача заключается в исследовании возможностей увеличения скорости передачи данных в системе UWB. Для этого в модели системы UWB, на основе Multiband OFDM, было сделано следующее:

– исследована зависимость вероятности битовой ошибки (BER) от скорости свёрточного кодирования (11/32, 1/2 и 5/8);

– исследована зависимость соотношения сигнал/шум для различных моделей среды распространения по отношению к общей вероятности битовой ошибки;

– исследовано влияние различных размеров OFDM символа (16, 12, 10 и 8 бит) на вероятность битовой ошибки.

Предложенные группой IEEE [8] четыре модели среды распространения и приёма сигнала можно охарактеризовать следующим образом:

– расстояние от 0 до 4 *м* без преград (CM1);

– расстояние от 0 до 4 *м* с преградами (CM2);

– расстояние от 4 до 10 *м* с преградами (CM3);

– расстояние от 4 до 10 *м* с преградами и среднеквадратическим отклонением задержки в 25 *нс* (CM4).

В первом случае, изменение вероятности битовой ошибки при скорости свёрточного кодирования 11/32, в условиях непрямой видимости на расстоянии до 10 *м* (CM4) наблюдается наилучшее значение вероятности битовой ошибки при заданном соотношении сигнал/шум. Вариант CM4 имеет запас мощности по отношению к наихудшему варианту CM2 примерно 1 *дБ*. Это связано с низкоскоростным избыточным кодированием, которое предотвращает влияние помех и разнородностей в канале передачи. Во втором случае, изменение вероятности битовой ошибки при скорости свёрточного кодирования уже 1/2, в случае прямой видимости на расстоянии до 4 *м* (CM1) наблюдается наилучшее значение вероятности битовой ошибки при заданном соотношении сигнал/шум. Вариант CM1 имеет запас мощности по отношению к наихудшему варианту CM4 уже 2 *дБ*. Здесь, в отличии от случая со свёрточным кодом 11/32, вносимой избыточности недостаточно для уменьшения вероятности ошибки при плохих условиях распространения сигнала. Это также касается и следующего случая со свёрточным кодом 5/8. И в случае третьего варианта, изменение вероятности битовой ошибки при скорости свёрточного кодирования 5/8, в условиях прямой видимости на расстоянии от 4 до 10 *м* (CM1) наблюдается наилучшее значение вероятности битовой ошибки при заданном соотношении сигнал/шум. Вариант CM1 имеет запас мощности по отношению к наихудшему варианту CM4 порядка 2 *дБ*.

После сравнительной характеристики различных условий передачи для заданных скоростей передачи, (рис. 9) приведём сравнительную характеристику вероятности битовой ошибки для различных скоростей передачи данных. Наибольший запас соотношения сигнал/шум наблюдается в случае со свёрточным кодом 11/32, в котором используется наибольшая избыточность, что даёт выигрыш по запасу мощности более 4 *дБ* при вероятности битовой ошибки $1 \cdot 10^{-6}$.

На рис. 10 отображена зависимость размера символа OFDM в PLPC кадре относительно вероятности битовой ошибки. Для наглядности было взято соотношению сигнал/шум равное 8 *дБ* и свёрточный код 5/8. В соответствии с технологией OFDM, были использованы четыре возможных варианта в битах: 16, 12, 10 и 8. При размер символа в 8 *бит* проявляется наи-

худший результат по вероятности битовой ошибки, и далее с последующим увеличением размера OFDM символа, начиная с размера в 10 *бит*, изменение вероятности битовой ошибки становится более пологим.

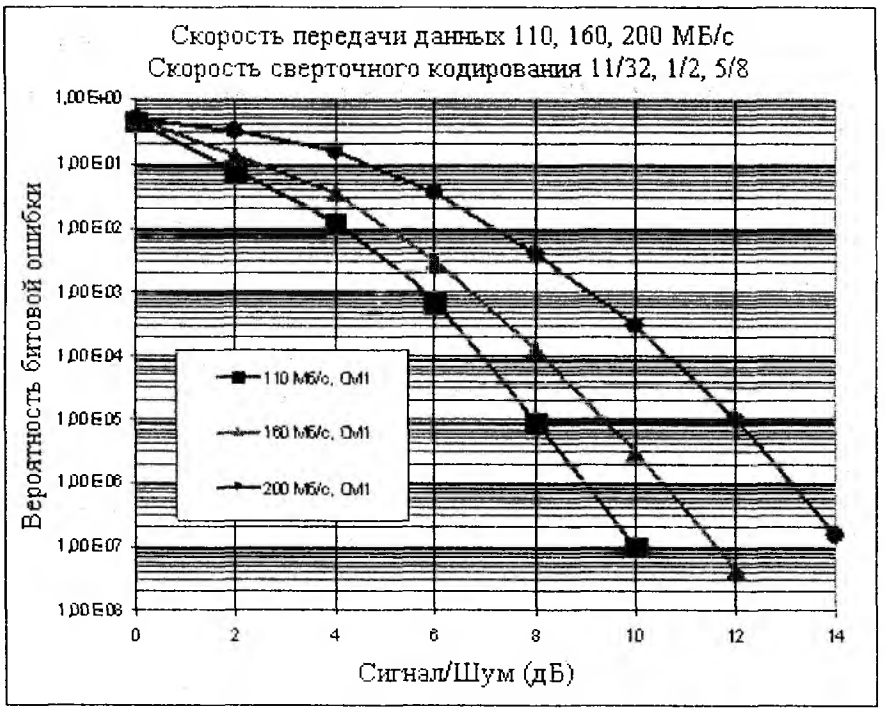


Рис. 9. Вероятности битовой ошибки для скоростей передачи данных 110, 160 и 200 *Мб/с* в канальной модели CM1

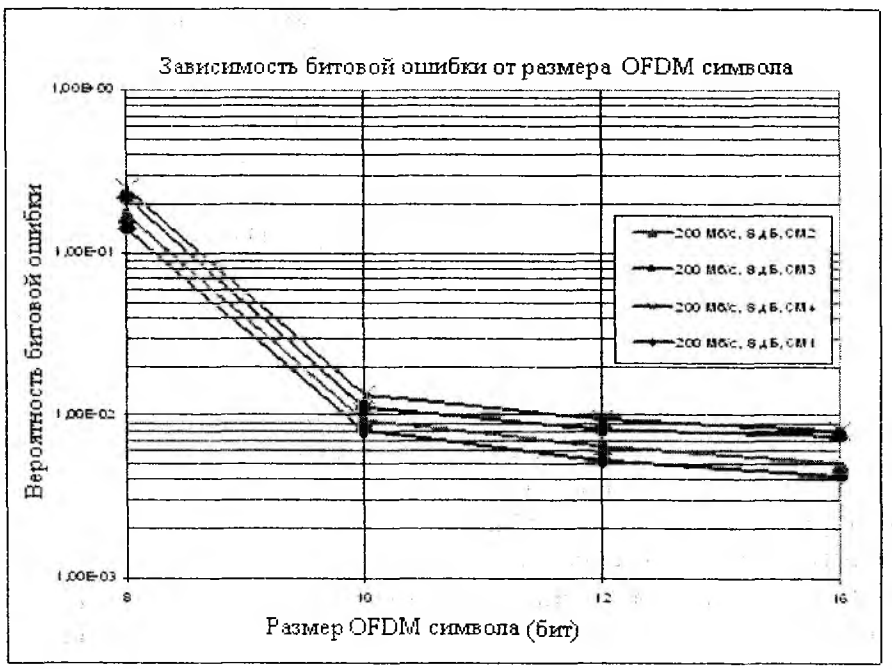


Рис. 10. Зависимость вероятности битовой ошибки от размера OFDM символа в битах

На основе проведённых исследований, на модели UWB, можно сделать следующие рекомендации:

– если приёмник и передатчик находятся в зоне прямой видимости, то лучше всего будет использование свёрточных кодов с малой избыточностью, например 5/8;

– когда же приёмник и передатчик не находятся в зоне прямой видимости, т.е. сигнал приходит переотраженный от различных объектов, то рекомендуется использовать свёрточный код 11/32, причем наибольший выигрыш по соотношению сигнал/шум получаем при наличии джиттера задержки сигнала порядка 25 *нс*;

– размер символа OFDM в технологии Mltiband OFDM не рекомендуется использовать равным 8 *бит*, в данном случае наблюдаются высокая вероятность ошибки. При увеличении символа до 10 *бит* вероятность ошибки уменьшается примерно в 80 раз, при дальнейшем увеличении OFDM символа вероятность ошибки уменьшается более медленно. Так при размере OFDM символа в 16 *бит* (наибольшее значение, предложенное в технологии Multiband OFDM), то вероятность ошибки уменьшается на 2 порядка, т.е. в 100 раз.

Заключение

Разработана модель приёмо-передатчика с использованием технологии UWB Multiband OFDM (стандарт IEEE 802.15.3a). Модель позволяет исследовать:

– зависимость вероятности битовой ошибки от скорости свёрточного кодирования;

– зависимость вероятности битовой ошибки от особенностей канала связи (прямая видимость сигнала или приём с переотражениями);

– зависимость вероятности битовой ошибки от расстояния между приёмником и передатчиком;

– зависимость вероятности битовой ошибки от размера символа OFDM.

На основании результатов моделирования даны рекомендации:

– если приёмник и передатчик находятся в зоне прямой видимости, то предлагается использовать свёрточный код с малой избыточностью, например 5/8;

– когда же приёмник и передатчик не находятся в зоне прямой видимости, т.е. сигнал приходит переотраженный от различных объектов, то рекомендуется использовать свёрточный код 11/32, причем наибольший выигрыш по соотношению сигнал/шум получается при наличии джиттера задержки сигнала порядка 25 *нс*;

– размер символа OFDM в технологии Multiband OFDM влияет на вероятность ошибки. При увеличении размера OFDM символа с 8 до 16 *бит* (наибольшее значение, предложенное в технологии Multiband OFDM), вероятность ошибки уменьшается на два порядка.

Список литературы: 1. *Ramjee, Prasad.* OFDM for wireless communications systems– Boston: Artech House, Inc., 2004. – 274 p. 2. *Intel Corporation.* Intel's Multi-band UWB PHY Proposal for IEEE 802.15.3a // IEEE 802.15.3a Working Group, submission, Mar. 2003. 3. *Рекомендация* IEEE 802.15.3a. Режим доступа: www/ URL: http://grouper.ieee.org/groups/802/15/ 4. *Ануфриев, И.Е., Смирнов, А.Б., Смирнова, Е.Н.* MATLAB 7. – СПб.: БХВ-Петербург, 2005. – 1104 с. 5. *Черных, И.В.* Simulink: среда создания инженерных приложений. – М. : Диалог-МИФИ, 2004. – 496 с. 6. *Черных, И.В.*Моделирование электротехнических устройств в MATLAB, SimPowerSystems и Simulink. – М. : Диалог-МИФИ, 2008. – 288 с. 7. *Bikramaditya, Das, Susmita Das.* Efficacy of Multiband OFDM Approach in High Data Rate Ultra WideBand WPAN Physical Layer Standard using Realistic Channel Models // International Journal of Computer Applications; Vol.2, No.2, May 2010 81–87 p. 8. *Foerster, J.* Channel Modeling Sub-Committee report final IEEE Doc.11.P802.15-02/490r1-SG3a, 2003.

Харьковский национальный
университет радиоэлектроники

Поступила в редколлегию 17.09.2012

УДК 621.391.23

И.В. КОВТУН, канд. техн. наук

ОСОБЕННОСТИ ПЕРЕДАЧИ ДАННЫХ ПО БЛОКАМ ТРАНКИНГОВОЙ СИСТЕМЫ СТАНДАРТА APCO 25

Постановка проблемы

В настоящее время процесс развертывания сетей транкинговой радиосвязи во всем мире характеризуется внедрением цифровых систем. Практически все ведущие мировые поставщики оборудования, системные интеграторы и операторы, а также многие крупные потребители услуг транкинговой радиосвязи переходят к цифровым системам. Основное соперничество на рынке стандартов, ориентированных не только на обычных корпоративных пользователей, но и на представителей правоохранительных органов и служб общественной безопасности, ведут TETRA, APCO 25 и TETRAPOL.

При сравнении цифровых стандартов целесообразно рассматривать те критерии, которые важны для потребителей услуг транкинговой радиосвязи, такие, как дальность связи и качество воспроизведения речевого сигнала.

Формулирование цели

Провести анализ основных перспективных стандартов цифровой транкинговой связи, дать им сравнительную характеристику и произвести оценку объема данных, которые передаются в системе APCO 25.

Сравнительная оценка различных стандартов транкинговой связи

Все показатели цифровых стандартов, которые можно рассматривать в качестве критериев для сравнения, условно можно разбить на две группы: эксплуатационно – технические и организационно – экономические [1].

Под эксплуатационно-техническими критериями понимается обобщенные технические показатели, которые определяются параметрами систем связи, такие, как дальность и оперативность связи, степень безопасности связи, спектральная эффективность, набор услуг связи, как стандартных, так и специфических. Каждый из этих критериев является комплексным, т.е., в свою очередь, складывается из нескольких показателей или зависит от определенного набора параметров.

К числу организационно экономических критериев можно отнести стоимостные показатели систем связи, возможности выделения ресурсов радиочастотного спектра и перспективы развития и распространения в мире каждого из стандартов. По сравнению с эксплуатационно-техническими показателями эти критерии имеют большую неоднозначность и гораздо более высокую степень субъективизма при их оценке. При этом организационно-экономические показатели в определенной степени зависят от технических, например на стоимостные показатели существенно влияют дальность связи и спектральная эффективность.

Рассматривая технические характеристики и функциональные возможности стандартов транкинговой связи TETRA, APCO 25 и TETRAPOL, можно отметить, что все стандарты имеют высокие (относительно данного класса систем подвижной радиосвязи) технические показатели. Стандарты позволяют использовать в своих системах дуплексные радиостанции. В средствах радиосвязи данных стандартов используются эффективные методы речепреобразования и помехоустойчивого кодирования информации. Все стандарты обеспечивают высокую оперативность связи и достаточную спектральную эффнеьтвность речепреобразования и помехоустойчивого кодирования информации. Все стандарты обеспечивают высокую оперативность связи и достаточную спектральную эффективность.

С технической точки зрения, основные различия между стандартами TETRA, с одной стороны, и APCO 25 и TETRAPOL – с другой, определяются методом разделения каналов

связи. Для стандарта TETRA – это многостанционный доступ с временным разделением каналов связи (МДЧР или FDMA). Рассмотрим, как это основное различие и другие технические параметры влияют на основные эксплуатационно-технические показатели.

Дальность связи

Дальность связи зависит от большого количества факторов, которые можно разделить на три основные группы:

- факторы, определяемые условиями применения средств связи (высота установки антенн, рельеф местности, помеховая обстановка и т.д);
- факторы, определяемые техническими параметрами, реализованными в аппаратуре связи (мощность передатчиков, чувствительность приемного тракта, коэффициент усиления антенны и т.д.);
- факторы, определяемые непосредственно заложенными в стандарте принципами построения каналов связи (ширина полосы канала связи, скорость информации в канале, способ модуляции сигнала, алгоритм речевого кодирования, методы помехоустойчивого кодирования).

Исходя из перечисленных групп корректно сравнивать стандарты цифровой транкинговой радиосвязи можно только по последней группе факторов, т.к. другие группы зависят или от условий эксплуатации, или определяются качеством производства радиосредств. Системы с FDMA обеспечивают большую дальность связи (при прочих равных параметрах) по сравнению с системами с TDMA. Это объясняется меньшей энергией сигнала на один бит информации. Известно, что энергия сигнала Ес определяется как

$$E_c = P_c \cdot T_c ,$$

где *Pc* – мощность сигнала; *Tc* – длительность сигнала.

Известно, что при уменьшении времени передачи одного информационного бита для цифровой системы пропорционально уменьшается энергия. Например, для систем TETRA, с четырьмя информационными каналами на одной физической частоте эквивалентная мощность на бит информации в четыре раза меньше, чем в системах с FDMA, что равносильно снижению дальности связи ориентировочно на 40 %.

Другим фактором, влияющим на снижение дальности связи в системах с TDMA по сравнению FDMA-системами, является устойчивость канала связи при многолучевом распространении сигнала, возникающим в условиях плотной городской застройки или холмистой местности из-за отражения сигнала от зданий и других преград, и приводящим к появлению радиоэха. Отраженный сигнал оказывает тем большее влияние, чем больше его отношение к длительности сигнала. Поэтому уменьшение длительности информационного бита в системах с TDMA ухудшает качество приема в условиях многолучевости.

Оперативность связи

Основным параметром, характеризующим оперативность связи, является время установления соединения (канала связи) между абонентами. Если рассматривать время установления канала связи в пределах зоны действия одной базовой станции, то все стандарты имеют близкие показатели, в пределах от 0,2 до 0,5 *с*. Однако преимущество стандартов, использующих FDMA (TETRAPOL, APCO 25), состоит в том, что минимальная длительность установления соединения сохраняется на более обширной территории, т.к. дальность связи для этих стандартов больше. Для абонентов сетей стандарта TETRA, в среднем, выше вероятность оказаться в разных зонах обслуживания. При этом вызов будет проходить через коммутатор, что неизбежно увеличит время установления соединения. Кроме того, существует опасность, что в зоне вызываемого абонента заняты все каналы ретранслятора, и даже в случае вытесняющего вызова потребуется время на разрыв одного из текущих соединений. Таким образом, в целом, можно сказать, что статистически время установления соединения для передачи речевых сообщений в сетях стандартов TETRAPOL и APCO 25 меньше, чем в стандарте TETRA.

Вместе с тем, большое значение в сетях подвижной радиосвязи приобретает скорость передачи данных, которая также является показателем оперативности связи. Для стандарта TETRA она может достигать 28,8 *Кбит/с* (при использование всех четырех временных интервалов для передачи массива данных). Для стандартов FDMA она в несколько раз меньше: для TETRAPOL – 8000 *бит/с*, для APCO 25 – 9600 *бит/с*.

Спектральная эффективность

Основным показателем спектральной эффективности системы связи является эффективная полоса частот на один речевой канал, определяющая какое количество каналов связи можно разместить в отведенной для развертывания сети связи фиксированной полосе частот. Из технических характеристик рассматриваемых стандартов видно, что по этому показателю TETRA имеет преимущество по сравнению со стандартами с частотным разделением каналов. Стандарт APCO 25 также декларирует эффективную полосу частот, равную 6,25 *кГц*, однако это будет достигнуто только во второй фазе реализации проекта.

Набор услуг связи

Рассматривая функциональные возможности представленных стандартов транкинговой связи, можно сказать, что они обеспечивают сравнимый уровень услуг связи. Все стандарты позволяют строить различные конфигурации сетей связи, обеспечивают разнообразные режимы передачи речи и данных, связь с телефонными сетями общего пользования (ТФОП) и фиксированными сетями. Стандарты позволяют использовать в своих системах дуплексные радиостанции. Некоторое преимущество имеют стандарты Tetrapol и TETRA, в которых реализованы режимы "двойного наблюдения" и открытого канала, крайне полезны для служб общественной безопасности. Однако, учитывая быстрое развитие стандартов и постоянное расширение функций систем связи, вполне возможно, что в скором времени такие же возможности будут предоставляться и APCO 25.

Организационно-экономические критерии

Наличие ресурсов радиочастотного спектра для развертывания системы радиосвязи является важнейшим критерием выбора той или оной системы. В данном случае наиболее перспективны стандарты, которые обеспечивают возможность построения сетей связи в наиболее широком диапазоне.

Системы TETRA теоретически обеспечивают возможность работы в очень широком диапазоне (150 – 900 *МГц)*. Вместе с тем, пока производители предлагают в основном оборудование, функционирующие только в диапазоне, выделенном в Европе для построения сетей TETRA – 380 – 385/390 – 395 и 410 – 470 *МГц*, хотя сейчас уже есть сведения о проектах систем в диапазоне 800 *МГц*.

Системы APCO 25 в соответствии с функциональными и техническими требованиями обеспечивают возможность работы в любом из диапазонов, отведенных для подвижной радиосвязи.

Стандарт TETRAPOL ограничивает верхнюю частоту своих систем на уровне 520 *МГц*. Реально большинство действующих систем используют диапазон 380 – 400 *МГц*.

Важным критерием сравнения стандартов является частотный ресурс, необходимый для развертывания сети связи с одинаковым количеством абонентов и одинаковой зоной радиопокрытия. С одной стороны, стандарт TETRA имеет лучшую спектральную эффективность, с другой – TETRAPOL и APCO 25 обеспечивают больший радиус зоны обслуживания базовой станции. Поэтому для систем TETRA меньшие ресурсы радиочастотного спектра будут требоваться для сетей радиосвязи с очень интенсивным трафиком, а преимущества TETRAPOL и APCO 25 будут проявляться для сетей связи с невысоким трафиком и широкой зоной охвата.

Экономическая эффективность

Сравнение экономической эффективности систем различных стандартов нельзя рассматривать в отрыве от категории системы подвижной радиосвязи. Для создания сетей связи с небольшой нагрузкой, широким территориальным охватом и числом каналов в пределах 10 более оптимальным вариантом (в т.ч. и по стоимости) является использование систем МДЧР, к которым относятся APCO 25 (Фаза I) и TETRAPOL. Это объясняется большим радиусом зон обслуживания систем МДЧР по сравнению с МДВР-системами. По оценкам, приведенным в техническом отчете стандарта TETRAPOL PAS, стоимость базового оборудования многозоновой сети радиосвязи, реализованного на основе МДВР по отношению к системе с частотным разделением каналов (при одинаковой стоимости единицы оборудования) будет на 30 – 50 % выше.

Однако для сетей связи с интенсивным трафиком и числом каналов в одной зоне более 15, предпочтительно использование систем с временным разделением каналов, к которым относятся TETRA. Следует отметить, что стандарт APCO 25 (Фаза II) будет обладать универсальностью, обеспечивая возможность строить системы, как с частотным, так и с временным разделением каналов.

Оценка объема данных, которые передаются в сети TETRA с коммутацией пакетов, была проведена в статье [2].

В данной работе рассмотрим особенности передачи данных в стандарте APCO 25.

Оценка объема передачи данных по блокам в стандарте APCO 25

В стандарте APCO речевая информация передается с помощью кадров по 180 мс, называемые логическими блоками данных (LDU). Общая длина каждого из них составляет 1728 *бит*. Группа из двух кадров создает суперкадр Информационная часть одного логического блока, которая передает саму речь, составляет 792 *бита*. Доля служебной информации составляет 936 *бит*. Передачи логических блоков предшествует преамбула, которая предназначена для начальной синхронизации передающей стороны и стороны, которая получает информацию, инициализации всех функций шифрования и передачи адресной информации. Ее размер составляет *Lпр* = 792 *бита*. В конце любого сообщения передается маркер конца сообщения, который может быть двух типов конструкции: простой и сложный. Размер простого маркера составляет *Lкон1* = 140 *бит*, а сложного – *Lкон2* = 288 *бит*.

Рассчитаем размер блока данных, в котором будет передаваться один суперкадр (два логических блока данных) по формуле (1) при использовании простого *Lпр* и сложного *Lсложн* маркера конца сообщения.

$$L = L_{пр} + 2 \cdot L_{LDU} + L_{сложн} \qquad (1)$$

Размер блока данных составит:

$$L_{пр} = 792 + 2 \cdot 1728 + 144 = 4392 \, бита,$$

$$L_{сложн} = 792 + 2 \cdot 1728 + 288 = 4700 \, бит.$$

Длина информационного блока, которую мы можем передать в одном суперкадре $L_{инф}$, составит

$$L_{инф} = 792 \cdot 2 = 1584 \, бита.$$

Аналогично проведем расчет для блоков данных, в которых будет передаваться два, три и четыре суперкадра. Полученные результаты занесем в таблицу.

Время передачи данных $t_{пер}$ рассчитывается по формуле (2).

$$t_{пер} = \frac{L}{V}, \qquad (2)$$

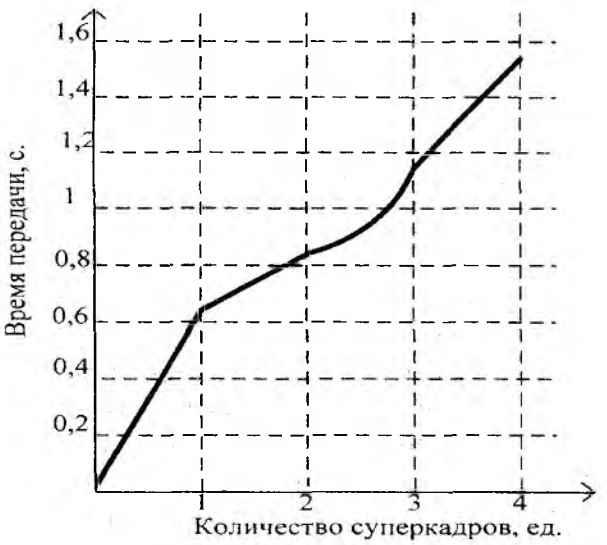
где L – длина информационного блока, *бит*; V – скорость передачи данных (для APCO 25 составляет *9,6 кбит/с*), *бит / с*.

Время передачи блока информации, состоящего из одного суперкадра с простым маркером конца сообщения, составляет $t_{пер1} = \frac{6120}{9600} = 0{,}6375 c.$

Рассчитаем время передачи и для блоков с двумя, тремя и четырьмя суперкадрами и занесем полученные данные в пятую колонку таблицы.

Количество суперкадров в блоке данных	$L_{инф}$, *бит*	$L_{пр}$, *бит*	$L_{сложн}$, *бит*	$t_{пер}$, с
1	1584	6120	4700	0,6375
2	3168	7848	8156	0,8175
3	4752	11304	11612	1,1775
4	6336	14760	15068	1,5375

На основе этих данных построим график зависимости времени передачи от количеств суперкадров в блоке, который передается. График представлен на рис. 1.



Выводы

Из рисунка видно, что при увеличении количества суперкадров в блоке информации, время передачи увеличивается. Поэтому необходимо уменьшить объем информации, чтобы она передавалась с помощью одного суперкадра, что, в свою очередь, уменьшит время передачи и увеличит скорость передачи.

Таким образом, для уменьшения времени передачи необходимо применять методы кодирования данных, позволяющие уменьшить объем данных, но при этом сохранять объем информации. Решением этой задачи неразрывно связано с применением методов и средств компактного представления или сжатия данных, позволяющих не только сократить их объем, но и минимизировать время передачи информации по каналам связи и формирования ее графической модели в устройствах отображения.

Применение методов и средств сжатия данных позволит увеличить также производительность и качество в целом всей информационной подсистемы и подсистемы хранения информации.

Список литературы: 1. *Громаков, Ю.А.* Стандарты и системы подвижной радиосвязи. – М. : Эко-Трендз Ко, 1997. – 238 с. 2. *Корольова, Н.А., Ковтун, І. В.* Оптимізація передачі відеоінформаційних потоків по пакетах в системах транкінгового зв'язку стандарту TETRA // Зб. наук. праць УкрДАЗ. - 2011. – № 127. – С. 39-43.

Украинская государственная академия железнодорожного транспорта

Поступила в редколлегию 07.09.2012

УДК 621.391.001

Ю.Г. ЛЕГА, д-р техн. наук

ЗАВАДОСТІЙКІСТЬ М-ПОЗИЦІЙНОГО АВТОКОРЕЛЯЦІЙНОГО ПРИЙМАЧА ШУМОВИХ СИГНАЛІВ В ГАУССОВОМУ КАНАЛІ

Вступ

В останні роки значно зросла зацікавленість фахівців до використання у телекомунікаційних системах шумоподібних (хаотичних) та шумових сигналів з розширеним спектром. Ця зацікавленість обумовлена можливістю передачі інформації з високими швидкостями та створенням мереж зв'язку з великою просторовою щільністю трафіку [1, 2]. Зростання вимог до захищеності інформації, обмеженості частотного ресурсу, розвиток елементної бази радіосистем з використанням досягнень мікроелектроніки та технологій цифрової обробки сигналів, дають змогу по новому підійти до проблем розробки ефективних систем зв'язку з шумовою носійною.

Постановка задачи

В [3] проаналізовано схему, що забезпечує підвищення швидкості передачі завдяки використання багатопозиційних шумових сигналів. Схема пристрою цієї системи зображена на рис. 1, де використані наступні позначення: Г – генератор шуму, τ – лінія затримки на час τ, $U_i(t)$, $i=\overline{1,m}$, – система псевдошумових ортогональних послідовностей, I_i – інтегратори.

Прикладом псевдошумової ортогональної системи $U_i(t)$, може слугувати послідовність Уолша, що для випадку, коли m=4, подана матрицею:

$$U = \|U_i(t)\| = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix},$$

де тактовий (системний) інтервал T розбитий на підінтервали довжиною $\Delta t = T/m$.

На виході передавача на одному тактовому інтервалі T сигнал має вигляд:

$$x(t) = \xi(t) + U_i(t)\xi(t-\tau), \tag{1}$$

де $i=\overline{1,m}$, – переданий інформаційний i-й символ m-позиційного коду.

На вході приймача спостерігається сигнал виду

$$y(t) = \xi(t) + U_i(t)\xi(t-\tau) + \eta(t), \tag{2}$$

де $\eta(t)$ – адитивна завада, яка додається в каналі зв'язку.

В схемі наведеного пристрою можна відмітити його апаратурну складність, що пов'язана з використанням m генераторів функцій $U_i(t)$ у передавальному та приймальному пристроях. Простішим структурним варіантом побудови m – позиційної системи є схема, що наведена на рис.2. В цій схемі використовуються m ліній затримки на інтервал часу τ_i, $i=\overline{1,m}$. Шумова складова сигналу $\xi(t-\tau_i)$ комутується комутатором по значенню символу α_i m-позиційного інформаційного коду, що сформований джерелом повідомлень.

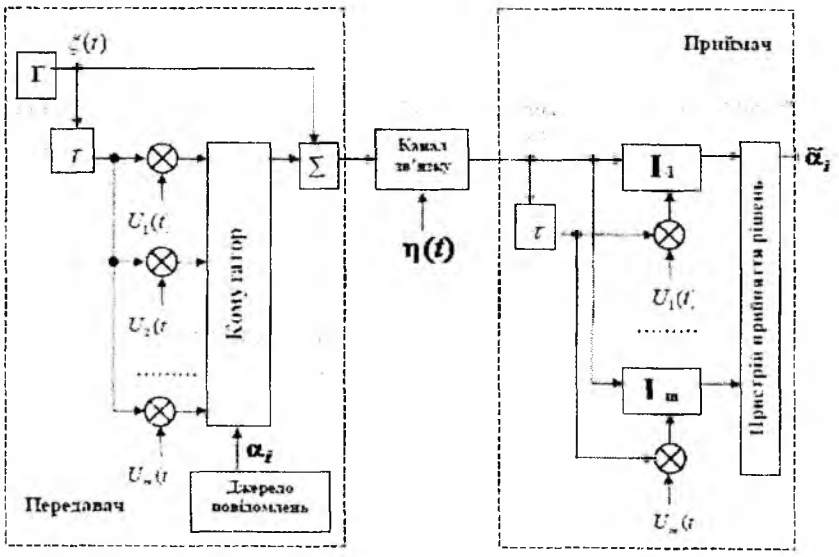


Рис.1

Метою даної роботи є теоретичний аналіз завадостійкості системи з m-позиційними шумовими сигналами, що зображена на рис. 2, при роботі по каналу з постійними параметрами в якому діє адитивна завада типу білого гауссового шуму.

Вирішення задачі

Результуючий сигнал передавача разом з завадою каналу зв'язку в цій системі має вигляд

$$y(t) = \xi(t) + \xi(t - \tau_i) + \eta(t), \quad i = \overline{1,m}. \tag{3}$$

Значення сигналу на k-у кореляційному пристрої, що спостерігається на виході інтегратору I_k при передачі сигналу по i-ї позиції, визначається величиною

$$\vartheta_k = \int_0^T (\xi(t) + \xi(t - \tau_i) + \eta(t))(\xi(t - \tau_k) + \xi(t - \tau_i - \tau_k) + \eta(t - \tau_k))dt. \tag{4}$$

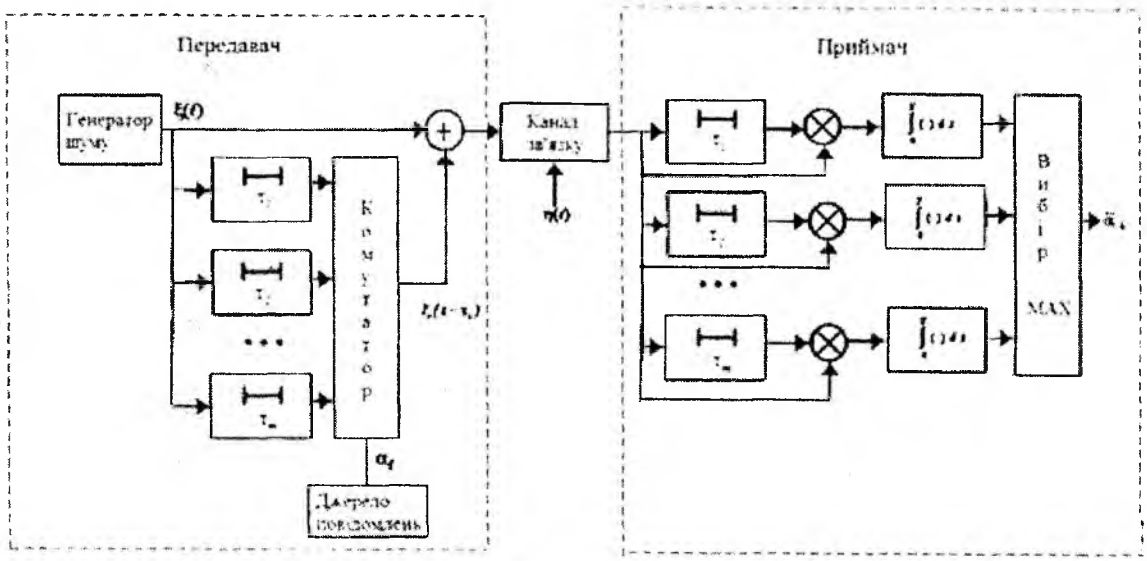


Рис. 2

Схема вибору максимального значення у приймачі фіксує прийнятий сигнал $\widetilde{\alpha}_k$ при виконанні умови

$$\vartheta_k = \max_i \vartheta_i . \tag{5}$$

Приймач фіксує вірне рішення про прийом символу $\widetilde{\alpha}_k$ у випадку, коли номер k-го кореляційного пристрою, співпадає зі значенням i-ї позиції сигналу α_i, що поступив для передачі. Як бачимо з (4), інтеграл від добутку $\xi(t-\tau_i)\xi(t-\tau_k)$ при значенні $k=i$ визначає складову у ϑ_i, що пропорційна енергії інформаційної складової сигналу, яка передана по каналу зв'язку. Всі інші складові характеризують додатки, що виникають за рахунок як дії адитивної завади у каналі зв'язку, так і обумовлені корисним сигналом, що створюють системну помилку.

Визначення ймовірності вірного прийому сигналу знайдемо з умовної сумісної щільності ймовірностей випадкових величин ϑ_k, $k=\overline{1,m}$ при передачі сигналу α_i. Для цього спочатку слід знайти значення початкових моментів випадкових величин ϑ_k, $k=\overline{1,m}$, заданих виразом (4).

Як математичну модель виходу генератора шуму $\xi(t)$ оберемо стаціонарний гауссовий дельта корельований центрований випадковий процес (гауссовий білий шум). Щільність розподілу одновимірного часового перерізу такого процесу має вигляд

$$f(\xi)=\frac{1}{\sqrt{2\pi D_\xi}}\exp(-\frac{\xi^2}{2D_\xi}), \tag{6}$$

де $D_\xi=\sigma_\xi^2$ – дисперсія шуму $\xi(t)$.

Розкриваючи дужки у виразі (4), можна виділити наступні доданки

$$\begin{aligned}\vartheta_k &= \varphi_{\xi,\xi}(t,t-\tau_k)+\varphi_{\xi,\xi}(t,t-\tau_i-\tau_k)+\varphi_{\xi,\eta}(t,t-\tau_k)+\\ &+\varphi_{\xi,\xi}(t-\tau_i,t-\tau_k)+\varphi_{\xi,\xi}(t-\tau_i,t-\tau_i-\tau_k)+\varphi_{\xi,\eta}(t-\tau_i,t-\tau_k)+\\ &+\varphi_{\eta,\xi}(t,t-\tau_k)+\varphi_{\eta,\xi}(t,t-\tau_i-\tau_k)+\varphi_{\eta,\eta}(t,t-\tau_k),\end{aligned} \tag{7}$$

де введено позначення $\varphi_{x,z}(a,b)=\int_0^T x(a)z(b)dt$, $x,z\in\{\xi,\eta\}$ – скалярний добуток процесів $x(a)$ та $z(b)$.

Обчислюючи значення початкових моментів випадкових величин, що присутні в (7), будемо вважати завади $\eta(t)$ центрованою (з нульовим значенням моменту першого порядку), а випадкові $\xi(t)$ і $\eta(t)$ стаціонарними та стаціонарно пов'язаними. При цих припущеннях, враховуючи лінійну властивість операції математичного сподівання та дельта корельованість випадкових процесів що розглядаються, маємо

$$m_1^{\varphi_{\xi,\xi}(t,t-\tau_k)}=M\int_0^T \xi(t)\xi(t-\tau_k)dt=\int_0^T M\{\xi(t)\}M\{\xi(t-\tau_k)\}dt=0.$$

Відповідно можна визначити величини:

$$\begin{aligned}&m_1^{\varphi_{\xi,\xi}(t,t-\tau_i-\tau_k)}=m_1^{\varphi_{\xi,\eta}(t,t-\tau_k)}=m_1^{\varphi_{\xi,\xi}(t-\tau_i,t-\tau_i-\tau_k)}=m_1^{\varphi_{\xi,\eta}(t-\tau_i,t-\tau_k)}=\\ &=m_1^{\varphi_{\eta,\xi}(t,t-\tau_k)}=m_1^{\varphi_{\eta,\xi}(t,t-\tau_i-\tau_k)}=m_1^{\varphi_{\eta,\eta}(t,t-\tau_k)}=0;\end{aligned} \tag{8}$$

$$m_1^{\varphi_{\xi,\xi}(t-\tau_i,t-\tau_k)} = \int_0^T M\{\xi(t-\tau_i)\xi(t-\tau_k)\}dt = \int_0^T \sigma_\xi^2 \delta(\tau_i-\tau_k)dt = \begin{cases} \sigma_\xi^2 T, & k=i \\ 0, & k\neq i, \end{cases} \tag{9}$$

де $\delta(\tau)$ – дельта-функція Дірака.

Узагальнюючи величини, представлені у виразах (8), (9), в (7) маємо

$$m_1^{\vartheta_k} = M\{\vartheta_k\} = \begin{cases} \gamma\sigma_\xi^2 T, & k=i, \\ 0, & k\neq i, \quad i=\overline{1,m}. \end{cases} \tag{10}$$

Обчислення другого початкового моменту $m_2^{\vartheta_k}$, $k=\overline{1,m}$, пов'язане з визначенням значення

$$\begin{aligned} m_2^{\vartheta_k} = M\{\vartheta_k^2\} = M\{[&\varphi_{\xi,\xi}(t,t-\tau_k)+\varphi_{\xi,\xi}(t,t-\tau_i-\tau_k)+\varphi_{\xi,\eta}(t,t-\tau_k)+ \\ &+\varphi_{\xi,\xi}(t-\tau_i,t-\tau_k)+\varphi_{\xi,\xi}(t-\tau_i,t-\tau_i-\tau_k)+\varphi_{\xi,\eta}(t-\tau_i,t-\tau_k)+ \\ &+\varphi_{\eta,\xi}(t,t-\tau_k)+\varphi_{\eta,\xi}(t,t-\tau_i-\tau_k)+\varphi_{\eta,\eta}(t,t-\tau_k)]^2\}. \end{aligned} \tag{11}$$

Визначаючи величини даних доданків вважаємо, що завада каналу зв'язку $\eta(t)$ має дисперсію σ_η^2. Обчислення доданків в (11) потребує розгляду інтегралів типу

$$\begin{aligned} m_2^{\varphi_{\xi,\xi}(t,t-\tau_k)} &= M\{\varphi_{\xi,\xi}^2(t,t-\tau_k)\} = M\{[\int_0^T \xi(t)\xi(t-\tau_k)dt]^2\} = \\ = M\{\int_0^T \xi(t)\xi(t-\tau_k)dt &\int_0^T \xi(x)\xi(x-\tau_k)dx\} = \int_0^T\int_0^T M\{\xi(t)\xi(t-\tau_k)\xi(x)\xi(x-\tau_k)\}dxdt. \end{aligned} \tag{12}$$

Момент четвертого порядку спільно гауссових центрованих випадкових величин $\xi(t), \xi(t-\tau_k), \xi(x), \xi(x-\tau_k)$ визначається наступною формулою [4]

$$\begin{aligned} &M\{\xi(t)\xi(t-\tau_k)\xi(x)\xi(x-\tau_k)\} = M\{\xi(t)\xi(t-\tau_k)\}M\{\xi(x)\xi(x-\tau_k)\}+ \\ &+M\{\xi(t)\xi(x)\}M\{\xi(t-\tau_k)\xi(x-\tau_k)\}+M\{\xi(t)\xi(x-\tau_k)\}M\{\xi(t-\tau_k)\xi(x)\}. \end{aligned} \tag{13}$$

Користуючись співвідношенням (13), з (12) одержимо

$$\begin{aligned} m_2^{\varphi_{\xi,\xi}(t,t-\tau_k)} &= \sigma_\xi^2\int_0^T\int_0^T M\{\xi(t)\xi(t-\tau_k)\}\delta(\tau_k)\}dxdt + \\ &+\sigma_\xi^2\int_0^T\int_0^T M\{\xi(t)\xi(x)\}\delta(t-x)dxdt + \\ &+\sigma_\xi^2\int_0^T\int_0^T M\{\xi(t)\xi(t-\tau_k)\}\delta(t-x-\tau_k)dxdt. \end{aligned}$$

Використовуючи фільтруючу властивість дельта-функції, отримаємо кінцеве значення

$$m_2^{\varphi_{\xi,\xi}(t,t-\tau_k)} = \sigma_\xi^4 T. \tag{14}$$

Враховуючи незалежність випадкових величин $\eta(t)$, $\xi(x)$ і фільтруючу властивість дельта-функції, можна знайти значення

$$m_2^{\varphi_{\xi,\eta}(t,t-\tau_k)} = \int_0^T \int_0^T M\{\eta(t)\eta(x)\}\sigma_\xi^2\delta(t-x)dxdt = \sigma_\eta^2\sigma_\xi^2 T\,. \qquad (15)$$

Подібним чином з (11) визначаємо наступні моменти

$$m_2^{\varphi_{\xi,\xi}(t,t-\tau_i-\tau_k)} = m_2^{\varphi_{\xi,\xi}(t-\tau_i,t-\tau_i-\tau_k)} = \sigma_\xi^4 T;\ m_2^{\varphi_{\eta,\eta}(t,t-\tau_k)} = \sigma_\eta^4 T;$$

$$m_2^{\varphi_{\xi,\eta}(t-\tau_i,t-\tau_k)} = m_2^{\varphi_{\eta,\xi}(t,t-\tau_k)} = m_2^{\varphi_{\eta,\xi}(t,t-\tau_i-\tau_k)} = \sigma_\xi^2\sigma_\eta^2 T\,. \qquad (16)$$

Значення величини

$$m_2^{\varphi_{\xi,\xi}(t-\tau_i,t-\tau_k)} = \begin{cases} \sigma_\xi^4(2T+T^2), & k=i, \\ \sigma_\xi^4 T, & k\neq i. \end{cases} \qquad (17)$$

Обчислення моментів подвоєних парних добутків елементів записаних у квадратних дужках виразу (11) приводить до наступного:

$$M\{2\varphi_{\xi,\xi}(t,t-\tau_k)\varphi_{\xi,\xi}(t,t-\tau_i-\tau_k)\} = 2\int_0^T \int_0^T M\{\xi(t)\xi(t-\tau_k)\xi(x)\xi(x-\tau_i-\tau_k)\}dtdx = 0.$$

Нульові значення мають і моменти всіх інших парних добутків доданків у виразі (11). Таким чином, узагальнюючи наведені вище результати у відповідних доданках в (11), можна записати

$$m_2^{\vartheta k} = M\{\vartheta_k^2\} = \begin{cases} [4\sigma_\xi^4 + 4\sigma_\xi^2\sigma_\mu^2 + \sigma_\eta^4)]T, & k\neq i,\quad i=\overline{1,m}; \\ [\sigma_\xi^4(5+T) + 4\sigma_\xi^2\sigma_\eta^2 + \sigma_\eta^4)]T, & k=i. \end{cases} \qquad (18)$$

Знаючи значення початкових моментів випадкових величин ϑ_i, $i=\overline{1,m}$, з (18) та (10) визначимо їх дисперсію

$$D_2^{\vartheta k} = m_2^{\vartheta k} - \left(m_1^{\vartheta k}\right)^2 = \begin{cases} [4\sigma_\xi^4 + 4\sigma_\xi^2\sigma_\eta^2 + \sigma_\eta^4]T, & k\neq i,\quad i=\overline{1,m}; \\ [5\sigma_\xi^4 + 4\sigma_\xi^2\sigma_\eta^2 + \sigma_\eta^4]T, & k=i. \end{cases}. \qquad (19)$$

Звернемо увагу на загальні властивості випадкових величин ϑ_k, $k=\overline{1,m}$, що випливають з аналізу їх складових. Як можна відмітити з розгляду схеми рис.2, кожна величина ϑ_k при значенні номеру субканалу приймача, що не співпадає з номером переданої позиції сигналу α_i, є інтегрований на інтервалі часу T добуток від незалежних перерізів в часі гауссових випадкових процесів. Тому такі випадкові величини ϑ_k залишаються гауссовими. Негауссовою є лише одна з величин ϑ_k, номер якої співпадає на даному тактовому інтервалі T з номером переданої позиції i. Закон щільності розподілу величини ϑ_i, яку можна розглядати як суму (що є наближенням інтегралу) квадратів гауссових випадкових величин (кожна складова підкоряється розподілу χ^2), при великій кількості складових в сумі наближається до гауссового. Тому, обмежуючись гауссовою апроксимацією і для складової ϑ_i, будемо в подальшому вважати, що кожна з величин ϑ_i, $i=\overline{1,m}$, має гауссовий розподіл.

Обчислення спільних моментів другого порядку випадкових величин ϑ_i, $i=\overline{1,m}$, можна виконати по розглянутій вище методиці. При цьому одержуємо наступні значення

$$M\{\vartheta_k\vartheta_i\} = 0,\quad k\neq i,\quad i=\overline{1,m}\,. \qquad (20)$$

Взаємна некорельованість гауссових випадкових величин ϑ_i, та ϑ_k, $k,i=\overline{1,m}$, $k\neq i$, вказує й на їх взаємну незалежність, що значно спрощує обчислення ймовірності помилки в розглянутій системі. Визначення ймовірності помилки виконаємо для випадку, коли сигнали передані по кожній з m позицій мають рівну дисперсію (потужність) й передаються з однаковою апріорною ймовірністю. Для цього скористаємося значенням умовної спільної щільності ймовірності $\omega(\vartheta_1,\vartheta_2,...\vartheta_m/\alpha_i)$ незалежних гауссових випадкових величин ϑ_i, $i=\overline{1,m}$, що має вигляд

$$\omega(\vartheta_1,\vartheta_2,...\vartheta_m/\alpha_i)=\prod_{k=1}^{m}\omega(\vartheta_k/\alpha_i). \tag{21}$$

де $\omega(\vartheta_k/\alpha_i)$ – умовна щільність імовірності випадкової величини ϑ_k при передачі сигналу i-ї позиції

$$\omega(\vartheta_k/\alpha_i)=\frac{1}{\sqrt{2\pi D_2^{\vartheta k}}}\exp\left(-\frac{(\vartheta_k-m_1^{\vartheta k})^2}{2D_2^{\vartheta k}}\right). \tag{22}$$

З урахуванням моментів і дисперсій, знайдених по виразам (10) і (19), одержимо

$$\omega(\vartheta_1,\vartheta_2,...\vartheta_m/\alpha_i)=\frac{1}{\sqrt{(2\pi)^m(D_2^{\vartheta_k})^{m-1}D_2^{\vartheta_i}}}\exp\left(-\frac{(\vartheta_i-m_1^{\vartheta_i})^2}{2D_2^{\vartheta_i}}-\frac{\sum\limits_{\substack{k=1,\\k\neq i}}^{m}(\vartheta_k)^2}{2D_2^{\vartheta_k}}\right). \tag{23}$$

Знаючи умовну щільність розподілу $\omega(\vartheta_1,\vartheta_2,...\vartheta_m/\alpha_i)$, можна знайти ймовірність правильного рішення при прийомі сигналу i-ї позиції:

$$P_{i,i}=\int_{-\infty}^{\infty}d\vartheta_i\int_{-\infty}^{\vartheta_i}...\int\omega(\vartheta_1,\vartheta_2,...\vartheta_m/\alpha_i)d\vartheta_1...d\vartheta_{i-1}d\vartheta_{i+1}...d\vartheta_m. \tag{24}$$

Результат інтегрування виразу (24) з підстановкою (23) запишемо як

$$P_{i,i}=\frac{1}{\sqrt{2\pi D_2^{\vartheta i}}}\int_{-\infty}^{\infty}\exp\left(-\frac{(x-m_1^{\vartheta i})^2}{2D_2^{\vartheta i}}\right)\hat{O}^{m-1}\left(x/\sqrt{D_2^{\vartheta k}}\right)dx, \tag{25}$$

де $\Phi(z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}\exp(-x^2/2)dx$ – інтеграл імовірності.

Отриманий вираз заміною змінною інтегрування може бути представлений у більш зручному для обчислень вигляду

$$P_{i,i}=\frac{1}{\sqrt{2\pi Q}}\int_{-\infty}^{\infty}\exp\left(-\frac{(z-G)^2}{2Q}\right)\hat{O}^{m-1}(z)dz, \tag{26}$$

де введені позначення: $G=m_1^{\vartheta i}/\sqrt{D_2^{\vartheta k}}$; $Q=D_2^{\vartheta i}/D_2^{\vartheta k}$.

Для апріорно рівно ймовірних сигналів повну ймовірність помилки, що визначає завадостійкість системи, визначимо зі співвідношення

$$P_{noм}=\sum_{i=1}^{m}(1-P_{i,i})p_i=\sum_{i=1}^{m}\frac{1}{m}(1-P_{i,i})=1-P_{i,i}, \tag{27}$$

де p_i – апріорна ймовірність передачі сигналу i-ї позиції.

Як видно з формул (27), (26), імовірність помилки в системі залежить від величини

$$G = m_1^{9i} / \sqrt{D_2^{9k}} = \sigma_\xi^2 \sqrt{T} / \sqrt{[4\sigma_\xi^4 + 4\sigma_\xi^2\sigma_\eta^2 + \sigma_\eta^4]} = \\ = \sqrt{qT/[4(1+q)+1/q]}, \tag{28}$$

де введене позначення $q = \sigma_\xi^2 / \sigma_\eta^2$ – перевищення дисперсії шумової складової сигналу над дисперсією завади каналу зв'язку.

Враховуючи введене позначення q, величина Q з (26) визначиться як

$$Q = \frac{5\sigma_\xi^4 + 4\sigma_\xi^2\sigma_\eta^2 + \sigma_\eta^4}{4\sigma_\xi^4 + 4\sigma_\xi^2\sigma_\eta^2 + \sigma_\eta^4} = 1 + (q/(1+2q))^2 . \tag{29}$$

З (29) відмітимо, що величина Q при варіації значення q змінюється мало і знаходиться в межах від 1 (при $q << 1$) до 1,25 (коли $q >> 1$). Тому завадостійкість системи від параметра q залежить не дуже сильно. Це проілюстровано на рис. 3, де представлена залежність $P_{пом}$ як функція від значення величини G, розрахована по (26) для різних значень q при кількості позицій m=2 та 8 (з позначеннями: —— – m =2, q =10; ····· - m =2, q =1; - - - - m =2, q =0,1; -·-· - m =8, q =1).

Залежність завадостійкості m – позиційної системи від величини G при різних значеннях кількості позицій m та при q =1, представлена на рис.4 (з позначеннями на рисунку: —— – m =2; ····· - m =8; - - - - m =16; -·-· - m =32).

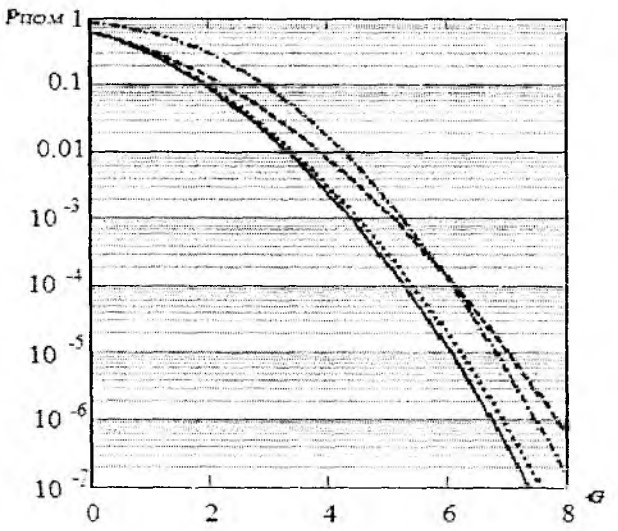


Рис. 3

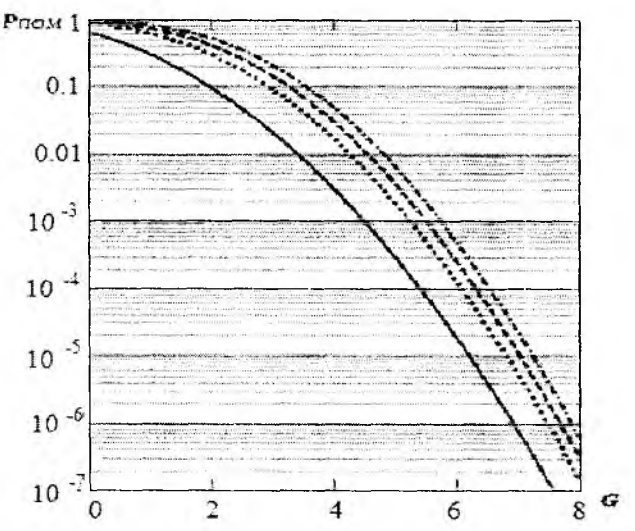


Рис. 4

Висновки

З даних, які наведені на рис. 3, можна заключити, що збільшення значення дісперсії шумової складової сигналу по відношенню до дисперсії завади каналу зв'язку (при сталому значенню G) покращує завадостійкість системи в діапазоні значень $q < 1$. При зростанні параметру $q > 1$ завадостійкість системи змінюється мало.

Збільшення числа позицій сигналу m веде до зниження завадостійкості системи (зростає ймовірність $P_{пом}$, рис.4). Але водночас зі зростанням значення m збільшується кількість

інформації ($\log_2 m$ біт – інформаційна ємність сигналу), що переноситься m – позиційним сигналом. Тому не значне погіршення завадостійкості системи при зростанні m, як видно з рис. 4, вказує на перспективність застосування систем з великою кількістю позицій шумового сигналу.

Список літератури: 1. *Лега, Ю.Г.* Системное проектирование средств связи с шумовыми сигналами. – К. : Наук. думка, 2000. – 304 с. 2. *Борисов,В.И., Зинчук, В.М., Лимарев, А.И. и др.* Помехозащищенность систем радиосвязи с расширением спектра сигналов модуляцией несущей псевдослучайной последовательностью. – М. : Радио и связь, 2003. – 640 с. 3. *Лега, Ю.Г., Первунінський, С.М., Гузнін,С.С.* Дослідження завадостійкості m-позиційного автокореляційного приймача шумових сигналів в каналі з адитивним білим гауссовим шумом // Вісник нац. універ. Львівська політехніка, серія «Радіоелектроніка та телекомунікації». – 2009. – № 645. – С. 167 – 173.

Черкаський державний
технологічний університет

Надійшла до редколегії 17.09.2012

УДК 621.397.7

С.А. ШЕЙКО, канд. техн. наук

ОСОБЕННОСТИ МЕЖКАДРОВОГО СЖАТИЯ ВИДЕОИНФОРМАЦИИ В УСТРОЙСТВАХ ВИДЕОНАБЛЮДЕНИЯ И ВИДЕОРЕГИСТРАЦИИ

Введение

Вопросы совершенствования методов сжатия видеоинформации, не смотря на стремительный рост производительности вычислительной техники, увеличение пропускной способности интернет-каналов и информационной ёмкости устройств хранения данных, остаются очень актуальными. Развитию методов сжатия способствует внедрение систем телевидения высокой чёткости, систем объемного изображения, а также специализированных прикладных телевизионных систем и устройств.

К наиболее востребованным инструментам на рынке прикладного телевидения относятся системы видеонаблюдения и видеорегистрации [1]. Современные системы видеонаблюдения постоянно совершенствуются и развиваются, используя всё большее количество видеокамер, всё чаще применяются ip-камеры и технологии Интернет, широкое распространение получают специализированные системы видеонаблюдения с борта радиоуправляемых носителей, например – беспилотных летательных аппаратов. Общей тенденцией является увеличение количества передаваемой и записываемой видеоинформации и возрастающие требования к качеству изображения.

Подавляющее большинство методов сжатия видеоинформации использует оценку и компенсацию движения для устранения временной избыточности видеоинформации, обусловленной схожестью соседних кадров. Это позволяет существенно сократить цифровой поток при прочих равных условиях. От точности компенсации движения непосредственно зависят степень сжатия и качество изображения. Поскольку сам метод оценки движения, как правило, не описывается в стандартах сжатия, к настоящему времени для этих целей разработано и применяется большое многообразие алгоритмов. Некоторые новые эффективные алгоритмы разрабатываются как комбинация уже существующих. В данной статье проводится анализ и экспериментальное исследование работы некоторых алгоритмов компенсации движения при сжатии видеоизображений в системах видеонаблюдения и видеорегистрации.

Анализ и экспериментальное исследование алгоритмов

Большинство видеокодеров построены на базе единой модели, включающей два основных пути прохождения данных: пространственное кодирование и предсказание с компенсацией движения [2]. На рис.1 показана структурная схема базового видеокодера.

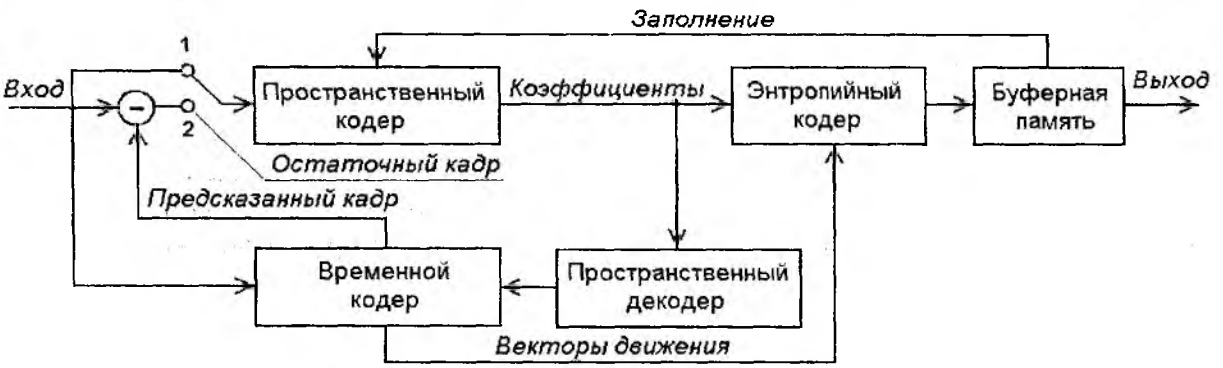


Рис. 1. Структурная схема базового кодера видеоданных

Кодер преобразовывает несжатую видеопоследовательность с помощью пространственного и временного кодеров, которые должны представить эту последовательность по возможно-

сти меньшим числом бит и с наиболее возможной точностью. Эти два требования чаще всего противоречат друг другу.

Процесс межкадрового кодирования состоит следующем. Несжатый видеокадр поступает на вход временного кодера, где сравнивается с опорным кадром. В качестве опорного чаще всего используется один из предшествующих кадров, реконструированных пространственным декодером. Реконструкция позволяет использовать в кодере точно такой же опорный кадр, что и в декодере, и тем самым избежать дрейфа ошибок при декодировании.

Функция временного кодера состоит в поиске в опорном кадре области, наиболее похожей на заданный блок текущего кадра. Смещение этой области относительно текущего блока фиксируется в виде вектора движения. Из найденных в опорном кадре наиболее похожих областей формируется предсказанный кадр, который вычитается из текущего кадра. Эта разность называется остаточным кадром и, по сути, является ошибкой предсказания. Коэффициенты пространственного кодера, а также вектора движения кодируются энтропийными кодерами, после чего поступают в буферную память. Большинство кодеров являются адаптивными, поскольку сигнал заполнения буферной памяти регулирует степень сжатия.

Главная задача межкадрового кодирования – уменьшение количества передаваемой информации за счёт снижения энергии остаточного кадра. Поэтому в большинстве случаев при поиске векторов движения применяются функции оценки подобия блоков, основанные на минимизации энергии разности текущего и предсказанного блоков [3]. Для блока размером $N \times N$ с координатами (x, y) оценочные функции вычисляются как

$$SAD(V_x, V_y) = \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \left| L_t(x+m, y+n) - L_{t-1}(x+V_x+m, y+V_y+n) \right|,$$
$$SSD(V_x, V_y) = \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \left(L_t(x+m, y+n) - L_{t-1}(x+V_x+m, y+V_y+n) \right)^2, \quad (1)$$

где (V_x, V_y) – значения вектора движения, L_t, L_{t-1} – значения яркости пикселей текущего и предыдущего кадров соответственно. Традиционно используется сумма абсолютных разностей (SAD – sum of absolute differences), поскольку вычисление данной функции не требует умножений.

Для большинства передаваемых сюжетов, снятых при большом отношении сигнал-шум, вектора соседних блоков мало отличаются, т.е. поле векторов движения достаточно гладкое. Поэтому для сокращения передаваемых данных дальнейшему кодированию и передаче подвергаются разностные вектора движения, имеющие компоненты

$$\Delta V_x^i = V_x^i - V_x^{i-1}; \; \Delta V_y^i = V_y^i - V_y^{i-1}, \quad (2)$$

где i – номер кодируемого блока.

В системах видеонаблюдения и видеорегистрации не редки случаи, когда съёмка производится в условиях малой освещённости. При уменьшении отношения сигнал-шум в видеосигнале увеличивается вероятность ошибочного определения векторов движения.

С целью выявления условий, при которых возникают ошибочные вектора движения, автором проведен ряд экспериментов, включавших видеосъёмку при различных условиях и определение полей векторов движения для отснятых видеопоследовательностей.

Один из примеров ошибочного определения векторов движения показан на рис.2. В данном случае видеосъёмка проводилась в офисном помещении при боковом освещении светильником с люминесцентной лампой мощностью 40 *Вт*. Диагональ ПЗС матрицы видеокамеры (1/4)", разрешение 0,8 *млн* пикселей, диафрагменное число объектива 1,6. Для уменьшения влияния прочих факторов съёмка осуществлялась видеокамерой без межкадрового сжатия (формат Motion JPEG), скорость цифрового потока 25 *Мбит/с*, размер кадра видеоизображения 720x576 пикселей. Поле векторов движения получено программным способом

по двум соседним кадрам методом полного перебора при минимизации функции (1). Размер блока составил 8x8 пикселей, размер окна поиска – 32x32 пикселя, точность поиска вектора движения – 1 пиксель.

Рис. 2. Пример поля «ошибочных» векторов движения

В данном эксперименте видеокамера неподвижна и движение в кадре отсутствует, тем не менее, найденные вектора движения существенно отличны от нуля. Вектора соседних блоков заметно отличаются, как по абсолютной величине, так и по направлению, поле векторов перестаёт быть гладким. Причём большее число векторов движения найдено в слабо освещённой части кадра, т.е. алгоритм, по сути, отслеживает изменение распределения шума от кадра к кадру. Такая ситуация наблюдалась в большинстве проведенных экспериментов и особенно характерна для съёмок достаточно однородных по своей структуре объектов.

Нахождение «ложных» векторов движения не приводит к искажениям изображения напрямую, найденные вектора соответствуют минимальной энергии предсказанного кадра. Однако данный фактор может снижать эффективность сжатия видеоинформации кодером. Применение оценочных функций SAD и SSD не позволяет учесть цифровой поток, требуемый для передачи векторов движения. При сжатии с большими коэффициентами, что весьма характерно для систем видеонаблюдения и видеорегистрации, а также в случае однородных видеоизображений, хорошо подвергающихся сжатию, цифровой поток для передачи векторов движения становится сравним с потоком, требуемым для передачи остаточных кадров. Так, при коэффициенте квантования 10 доля векторов движения составляет около 10 % от общего цифрового потока, при параметре 20 – около 20 %, при параметре 40 – около 35 % [4]. Поскольку сигнал заполнения буферной памяти регулирует поток данных, управляя коэффициентом сжатия, то при увеличении потока векторов движения следует ожидать увеличения потерь в пространственном кодере.

Описанной ситуации можно избежать, применив при поиске векторов движения оценочную функцию, учитывающую не только энергию остаточного кадра, но и битовые затраты на кодирование векторов движения. Одна из таких функций предложена в [5] в виде

$$Cost(V_x^i, V_y^i) = SAD(V_x, V_y) + \lambda \cdot r\left(V_x^i, V_y^i\right), \tag{3}$$

где $r(V_x^i, V_y^i)$ – количество бит, необходимое для передачи составляющих векторов движения текущего блока, λ – множитель, учитывающий степень компрессии.

Для минимизации выражения (3) нужно знать значение λ. Один из вариантов второго слагаемого и оценочной функции предложен в работе [6] в виде

$$Cost(V_x^i, V_y^i) = SAD(V_x, V_y) + k_t l\left(\left|V_x^i - V_x^{i-1}\right| + \left|V_y^i - V_y^{i-1}\right|\right), \quad (4)$$

где k_t – коэффициент, учитывающий степень компрессии текущего кадра. $l = 0{,}3$ – постоянный коэффициент, найденный в [6] экспериментально.

Второе слагаемое в (4) учитывает, что кодер передаёт разность составляющих векторов движения текущего и предыдущего блоков. При некоторых допущениях величина k_t может быть представлена в следующем виде:

$$k_t = \left(\sum_{i=1}^{M} S_{t-1}^i\right) / \left(\sum_{i=1}^{M} SAD_{t-1}^i\right), \quad (5)$$

где M – количество блоков в кадре; S_{t-1}^i – число бит для кодирования i-го блока в $(t-1)$-м кадре; SAD_{t-1}^i – значения функции (1) для i-го блока в $(t-1)$-м кадре.

Для оценки эффективности использования оценочной функции (4) при межкадровом сжатии в системах видеонаблюдения и видеорегистрации автором проведен ряд экспериментов по нахождению полей векторов движения для отснятых и описанных ранее видеопоследовательностей.

На рис.3 показано поле векторов движения, найденное для тех же условий, что и поле на рис.2, однако подобие блоков в сравниваемых кадрах оценивалось согласно выражениям (4) и (5).

Рис. 3. Поле векторов движения

Как видно из рисунка, найденное поле векторов движения состоит преимущественно из векторов малой величины и существенно более гладкое – вектора соседних блоков мало отличаются, что уменьшит битовые затраты на их передачу.

Проведена количественная оценка уменьшения скорости цифрового потока при сжатии с использованием оценочной функции (4). Эксперимент проведен для трёх видеопоследовательностей: первая и вторая получены при съёмке в офисном помещении и имеют отношение сигнал-шум 24 и 30 *дБ* соответственно, третья – при съёмке на улице пасмурным днём, отношение сигнал-шум 38 *дБ*. Значения отношения сигнал-шум рассчитаны по яркостной компоненте согласно выражению

$$PSNR = 10 \lg \frac{255^2 \cdot X \cdot Y}{\sum_{x=1}^{X} \sum_{y=1}^{Y} (L(x,y) - \overline{L}(x,y))^2}. \quad (6)$$

где $X = 720$, $Y = 576$ – размеры кадра в пикселях, $L(x, y)$ – отсчёты яркости произвольного кадра видеопоследовательности, $\overline{L}(x, y)$ – отсчёты яркости, усреднённые по ста кадрам.

В вычислении выражения (6) участвовали только те области изображения, в которых отсутствовало движение. Во всех трёх видеопоследовательностях движущиеся объекты занимали не более 20 % площади кадра.

Сжатие производилось по алгоритму MPEG-2 с различными коэффициентами квантования, поиск векторов движения осуществлялся методом полного перебора. Оценивалась относительная разница D в размерах видеопоследовательностей, сжатых с использованием оценочных функций *Cost* (4) и *SAD* (1):

$$D = \frac{S_{Cost} - S_{SAD}}{S_{SAD}} \cdot 100\%. \qquad (7)$$

График зависимости величины D от коэффициента квантования f для трёх тестовых видеопоследовательностей показан на рис.4.

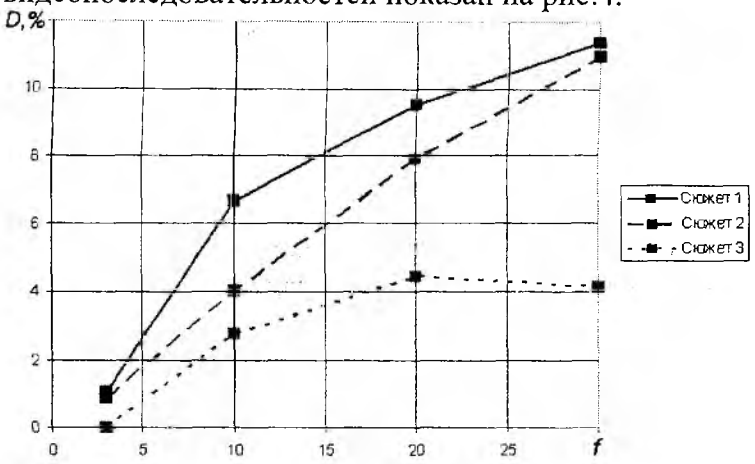


Рис. 4. Выигрыш в размере видеопоследовательностей, сжатых с использованием оценочных функций Cost и SAD

Анализ полученных данных показывает, что по сравнению с функцией *SAD* использование оценочной функции *Cost* даёт выигрыш при коэффициенте квантования 30 около 11 %. Наибольший выигрыш получается при сжатии видеопоследовательностей 1 и 2, имеющих меньшее отношение сигнал-шум, а также большее количество крупных однородных деталей, легче подвергающихся сжатию. Видеопоследовательность 3 имеет заметно лучшее отношение сигнал-шум, а также содержит достаточно большое количество мелких деталей, что позволило получить выигрыш около 4 % при коэффициенте квантования 15 и более.

Также было уточнено значение коэффициента l в выражении (4). При $l = 0{,}34$ для тестируемых видеопоследовательностей получены наилучшие результаты.

В проведенном эксперименте использовался метод полного перебора как эталонный, поскольку в заданном окне поиска он позволяет найти абсолютный минимум оценочной функции. Однако такой метод поиска обладает высокой вычислительной сложностью, так как для определения наилучшего вектора движения при окне поиска размером $N \times N$ необходимо вычислить оценочную функцию $(N+1)^2$ раз.

Для уменьшения количества операций предложено большое количество методов, основанных на ограниченном рассмотрении возможного числа кандидатов на вектора движения [3]. Важным свойством каждого такого алгоритма является его устойчивость к воздействию шума.

Рассмотрим свойства сигнала и шума, характерные для систем видеонаблюдения и видеорегистрации. На рис.5 показана частотно-контрастная характеристика (ЧКХ) использованной в экспериментах видеокамеры, снятая по штриховой мире в центре испытательной таблицы ТИТ-0249 для двух значений диафрагменного числа F при выключенном корректоре чёткости.

Завал ЧКХ видеокамеры на верхних частотах обусловлен рядом причин, главными среди которых являются конечная разрешающая способность объектива и матричного фотоприёмника [7]. Особенно заметна потеря чёткости при максимально открытой диафрагме

($F = 1{,}6$), что характерно для съёмок при малой освещённости. Учитывая также, что в пространственном спектре большинства изображений основная энергия сосредоточена в низкочастотной области, а энергетический спектр флуктуационного видеошума равномерный, можно сделать вывод, что отношение сигнал-шум в системах видеонаблюдения, работающих без специального освещения, на верхних частотах будет существенно меньше, чем на нижних. Следует полагать, что именно этими причинами обусловлено наличие «ложных» векторов движения в однородных областях без мелких деталей, наблюдаемое во всех проведенных экспериментах. Коррекция ЧКХ видеокамеры в области верхних частот не улучшает ситуацию, поскольку не увеличивает отношение сигнал-шум.

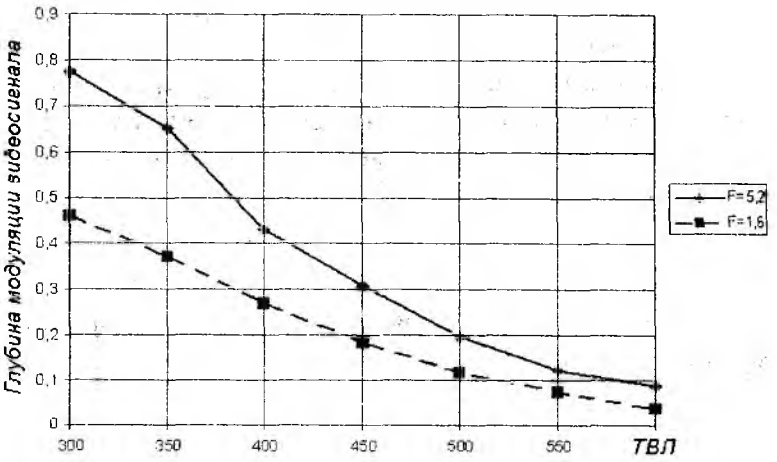


Рис. 5. Частотно-контрастная характеристика видеокамеры

Таким образом, для уменьшения ошибок в определении векторов движения целесообразно применять метод, в котором первоначально проводится поиск движения для низкочастотных компонентов изображения, имеющих большее отношение сигнал-шум, а затем произвести уточнение вектора с использованием высокочастотных компонентов спектра. Возникновение ошибок при таком методе поиска возможно лишь на этапе уточнения, что существенно уменьшает общую ошибку вычисления вектора движения.

Одним из вариантов иерархических алгоритмов является «средняя пирамида» [8]. В этом методе с помощью фильтра нижних частот из исходного кадра формируется кадр с вдвое меньшим разрешением по горизонтали и вертикали. Полученный кадр снова подвергается фильтрации, такая процедура повторяется P раз, где P – параметр алгоритма. Общее количество уровней иерархии составляет $P+1$, исходный и опорный кадры соответствуют нулевому уровню. Передискретизирующий фильтр может быть описан выражением [8]:

$$L_p(x,y) = \sum_{u=0}^{1}\sum_{v=0}^{1} L_{p-1}(2x+u;2y+v), \qquad (8)$$

где $L_p(x,y)$ – яркость пикселя с координатами (x,y) кадра уровня p.

Для оценки эффективности использования иерархического метода поиска векторов движения в видеопоследовательностях с малым отношением сигнал-шум проведен эксперимент на описанных выше видеосюжетах. Поиск выполнялся при параметре алгоритма $P = 2$, формировались три уровня иерархии. Один пиксель уровня 2 соответствует блоку 4x4 на нулевом уровне и блоку 2x2 на первом уровне иерархии. На верхнем уровне пирамиды использован трёхшаговый поиск, на остальных уровнях – полный поиск в небольших окнах. Вектору движения, найденному на уровне p, присваивался вес 2^p.

Об эффективности иерархического метода можно судить по среднему модулю отклонения компонентов найденных векторов движения $(V_{x\grave{e}}, V_{y\grave{e}})$ относительно найденных методом полного перебора $(V_{x\ddot{i}}, V_{y\ddot{i}})$:

$$\Delta = \left(\frac{|V_{x\ddot{i}} - V_{x\grave{e}}|}{V_{x\ddot{i}}} + \frac{|V_{y\ddot{i}} - V_{y\grave{e}}|}{V_{y\ddot{i}}} \right) \cdot 100\%. \qquad (9$$

В проведенном эксперименте относительное отклонение Δ не превысило 4 %, что подтверждает целесообразность применения иерархического метода при поиске векторов движения в кодерах систем видеонаблюдения.

Выводы

Результаты проведенных экспериментов демонстрируют, что при сжатии видеоинформации в системах видеонаблюдения и видеорегистрации для поиска векторов движения эффективно использовать оценочные функции, учитывающие битовые затраты на передачу векторов движения. Это позволяет снизить скорость цифрового потока до 11 % при параметре квантования 30. Наибольший выигрыш наблюдается при сжатии видеоизображений с невысоким отношением сигнал-шум и с малым количеством мелких деталей. При сжатии с бóльшими параметрами квантования выигрыш в скорости увеличивается, поскольку в этом режиме количество бит, требуемое для кодирования векторов движения, сравнимо с количеством бит для кодирования разностного блока.

Также показано, что для устойчивости алгоритма поиска векторов движения к воздействию шума в системах видеонаблюдения целесообразно применять иерархический метод поиска. Относительная ошибка поиска вектора движения при этом по сравнению с методом полного перебора не превышает 4 %.

Список литературы: 1. *Кругль, Г.* Профессиональное видеонаблюдение. Практика и технологии аналогового и цифрового CCTV / Г. Кругль ; пер. с англ. – М. : Секьюрити Фокус, 2010. – 640 с. 2. *Ричардсон, Я.* Видеокодирование. H264 и MPEG-4 – стандарты нового поколения / Я. Ричардсон ; пер. с англ. – М.: Техносфера, 2005. 368 с. 3. *Кубасов Д.* Обзор методов компенсации движения. [Электронный ресурс] / Кубасов Д., Ватолин Д. // Компьютерная графика и мультимедиа, Сетевой журнал. 2005. №3(2). URL: http://cgm.computergraphics.ru/content/view/76. 4. *Беляев, Е.А.* Алгоритмы оценки движения в задачах сжатия видеоинформации на низких битовых скоростях / Беляев Е.А., Тюрликов А.М. // Компьютерная оптика. – 2008. – Т. 32, №4. – С. 403–413. 5. *Kossentini, F.* Predictive RD Optimized Motion Estimation for Very Low Bit-Rate Video Coding / F. Kossentini, Y.-W. Lee, M.J.T. Smith and other // IEEE Journal On Selected Areas in Communications. – 1997. – Vol. 15, N.9. – P. 1752–1763. 6. *Потапов, П.В.* Оценочная функция для поиска векторов похожести при сжатии видеоданных / Потапов П.В., Кориков А.М. // Доклады ТУСУРа. – 2008. – № 2(18), ч.2. – С.87-90. 7. *Куликов, А.Н.* Реальная разрешающая способность телевизионной камеры. [Электронный ресурс] // Специальная техника. 2002, № 2. URL: http://www.ess.ru/sites/default/files/files/articles/2002/02/2002_02_04.pdf (дата обращения 15.06.2012). 8. *Kwon Moon Nam.* A Fast Hierarchical Motion Vector Estimation Algorithm Using Mean Pyramid / Kwon Moon Nam, Joon-Seek Kim, Rae-Hong Park // IEEE Transactions on Circuits and Systems for Video technology. – 1995. – Vol. 5, N.4. – P. 344–351.

Харьковский национальный
университет радиоэлектроники

Поступила в редколлегию 08.09.2012

УДК 681.326

В.Г.КОТУХ, канд. техн. наук, М.А. МИРОШНИК, канд. техн. наук,
С.Н. СЕЛЕВКО, канд. техн. наук

МЕТОДЫ ПЛАНИРОВАНИЯ РЕСУРСОВ В РАСПРЕДЕЛЕННЫХ КОМПЬЮТЕРНЫХ СИСТЕМАХ

Введение

Анализ современного состояния и тенденций развития компьютерных систем (КС) с использованием новых информационных и субмикронных технологий определяет новые требования к производительности и отказоустойчивости мультипроцессорных систем.

С точки зрения отказоустойчивости, функционирование КС представляет собой последовательность чередующихся интервалов работоспособного состояния и восстановления работоспособности путем реконфигурации структуры. Основанием процедуры реконфигурации КС и восстановления работоспособности является идентификация состояния системы и обнаружение неисправных модулей в процессе профилактического обслуживания и тестового диагностирования внешними или встроенными средствами диагностирования. С ростом сложности КС возрастает время и стоимость процедуры тестового диагностирования, что является причиной уменьшения производительности всей системы и эффективности её использования. Идеальным решением этой проблемы является использование методов функционального (онлайнового) диагностирования КС, когда в качестве тестов используются рабочие воздействия, поступающие в процессе функционирования системы. При этом исключаются или сводятся к минимуму интервалы времени профилактических тестовых проверок работоспособности системы.

Использование в составе КС процессорных модулей с интеллектуальными свойствами (IP-modules) определяет необходимость построения внутренней инфраструктуры КС, управляющей процессом оптимизации производительности и функциональной эффективности КС, а также онлайновым процессом тестового диагностирования, локализации неисправных модулей и реконфигурации её структуры. В ряде исследований показано, что потребление энергии в режиме тестового диагностирования в полтора раза превышает энергию, потребляемую в процессе нормального диагностирования, и этот разрыв имеет тенденцию увеличиваться с ростом сложности компьютерных сетей.

Для построения диагностической инфраструктуры с интеллектуальными свойствами (ДИ-ИС) можно использовать диагностические процессоры (модули), которые осуществляют онлайновое тестовое диагностирование КС. Однако при этом необходимо управлять процессом переключения процессорных модулей КС с режима функционирования в режим тестового диагностирования с учетом ограничений на допустимое время выполнения функциональных программ управления и потребляемой энергии на выполнение тестового диагностирования.

Известно, что для многих технологических процессов характерны непрерывный процесс протекания в пределах суточных интервалов и длительные циклы повторяемости (до нескольких десятков лет). Кроме того, к ним предъявляются жесткие требования, регламентируемые отраслевыми и государственными стандартами, связанные с высоким уровнем ответственности и серьёзными техногенными рисками. Поэтому диагностическое обслуживание подобных систем может осуществляться либо в режимах микродиагностики (в промежутках между вычислениями управляющих воздействий), либо по окончании технологического процесса. В настоящей статье поставлена задача разработки концепции диагностической инфраструктуры, обеспечивающей совмещение процессов функционального и тестового диагностирования КС в процессе управления сложными динамическими объектами, разработки программно-технических средств реализации отказоустойчивых КС.

GRID-технологии

Технологии *GRID* используются для создания географически распределенной вычислительной инфраструктуры, объединяющей ресурсы различных типов с коллективным доступом к ним в рамках виртуальных организаций, состоящих из предприятий и специалистов, совместно использующих эти ресурсы. Развитие и внедрение технологий *GRID* носят стратегический характер. В ближайшей перспективе эти технологии позволят создать принципиально новый вычислительный инструмент для развития высоких технологий в различных сферах человеческой деятельности. Идейной основой технологии *GRID* является объединение ресурсов путем создания компьютерной инфраструктуры нового типа, обеспечивающей глобальную интеграцию информационных и вычислительных ресурсов на основе сетевых технологий и специального программного обеспечения промежуточного уровня (между базовым и прикладным ПО), а также набора стандартизованных служб для обеспечения надежного совместного доступа к географически распределенным информационным и вычислительным ресурсам: отдельным компьютерам, кластерам, хранилищам информации и сетям. Отметим высокую актуальность исследований, проводимых в данном направлении, которая подтверждается многочисленными научными работами отечественных и зарубежных авторов [1 – 9].

Эффективность функционирования *GRID*-систем как среды с коллективной формой обслуживания пользователей определяется, в первую очередь, согласованностью распределения имеющихся ресурсов, которое должно происходить автоматически, опираясь на планирование вычислительных процессов в *GRID* в целом. Поэтому одной из ключевых функций, требуемых от программного обеспечения *GRID*-систем, является функция диспетчеризации, которая обеспечивает распределение ресурсов из общего ресурсного пула между заданиями, доставку программ и данных для их последующего выполнения [6 – 9]. В архитектуре *GRID*-системы функция диспетчеризации реализуется специальными программными службами, обеспечивающими такой уровень интеграции распределённых ресурсов, при котором *GRID*-система представляется в виде единой операционной среды обработки запросов (заданий, приложений, требований). Совокупность таких служб составляет систему диспетчеризации. Большинство существующих на сегодня систем диспетчеризации предназначено для обслуживания вычислительных *GRID*-систем, состоящих из кластеров (кластеризованных ресурсов) – традиционной формы организации распределённых вычислительных ресурсов. Используемым на практике системам диспетчеризации присущи довольно жёсткие ограничения на их применение: они не способны исключить такие нежелательные эффекты, как задержки времени обработки заданий, задержка обработки в ситуациях, когда имеются простаивающие вычислительные ресурсы, неравномерность загрузки ресурсов кластерах *GRID*-систем. Диспетчеризация включает в себя систему планирования [8, 9], которая обеспечивает накопление и затем гарантированное выделение ресурсов на кластерах (коаллокацию ресурсов): это предотвращает «зависание» заданий, которое является следствием фрагментации ресурсного пула.

Одной из основных задач для эффективного использования ресурсов *GRID* является задача планирования ресурсов для назначения на них заданий внешних и внутренних потоков *GRID*. Отметим, что использование, например, централизованных методов планирования, несмотря на их достаточно высокую популярность в современных промышленных *GRID*, не всегда позволяет эффективно решать проблемы загруженности имеющихся ресурсов. С этой точки зрения более производительными являются иерархические схемы построения *GRID*-систем. Основная идея, положенная в основу этих исследований, заключается в том, что на первом уровне планирования определяются те задания, для которых наиболее оптимальным образом определяются ресурсы, на которые они распределяются для решения (при этом используется математические модели упаковки в контейнеры). Результатом практического использования такого подхода является эмуляция в реально работающую систему, которая показала лучшие результаты обеспечения равномерной загруженности используе-

мых ресурсов. Вместе с тем, разработка формализованных моделей первого уровня планирования иерархической *GRID*-системы, на котором решаются и задачи метапланирования, на наш взгляд, требует проведения теоретических и практических исследований.

Цель данной работы – разработка нового подхода к планированию распределения ресурсов в гетерогенных *GRID*-системах, позволяющего повысить качество обслуживания заданий пользователей, а именно планирования, позволяющего обеспечить равномерную загрузку вычислительных ресурсов, повысить оперативность обработки поступающих на выполнение заданий, уменьшить общее время решения заданий на основе математической модели задачи о наименьшем покрытии и предложить вариант практической реализации метода. Реализацию нового подхода предполагается осуществить на основе имитационной модели, в основе которой лежит двухуровневая архитектура *GRID*.

Базовая конструкция и принципы работы модели

Для описания конструкции и работы предлагаемой модели используем следующие понятия и определения. В качестве базовой архитектуры модели выбирается архитектура вычислительной *GRID*-системы, использующая иерархическую структуру: на первом уровне находится система планирования ресурсов всей *GRID*-системы (планировщик первого уровня), на втором – локальном – уровне осуществляется локальное планирование заданий на ресурсе (локальный планировщик), предназначенное для планирования решения назначенных на данный ресурс заданий. Такая система, в соответствии с классификацией *GRID*-систем, называется двухуровневой иерархической системой – с планировщиками верхнего уровня и локального уровней планирования заданий [3 – 8].

В двухуровневой архитектуре предлагается использовать следующие компоненты:

- глобальная очередь, которая формируется по мере осуществления подписки пользователями своих заданий на их решение в *GRID*; глобальная очередь характеризуется размером L_{GlQuer}, количеством задач, поступивших на их решение в *GRID* за время формирования очереди T_{GlQuer};
- промежуточный пул заданий, выбранных из глобальной очереди, имеющий размер L_{pool} ($L_{pool} \leq L_{GlQuer}$). Размер пула может определяться, например, периодичностью процесса планирования: днями, неделями и т.д.;
- пакет заданий, формируемый как результат обработки заданий пула на основе решения задачи о наименьшем покрытии, имеющий определенный размер L_{buffer} и представляющий собой очередь на локальный ресурс.
- ресурсы R_i , представляющие собой кластеры *GRID* системы, на которых будут решаться спланированные планировщиком верхнего уровня задания.

Последовательность обработки заданий, находящихся в глобальной очереди, следующая:

- задание выбирается из глобальной очереди и помещается в пул по определенному принципу (приоритету, стоимости и т.д.);
- планировщик верхнего уровня в соответствии с режимом (политикой) обслуживания очереди планирует задания на свободные и доступные на данный момент времени ресурсы на основе решения задачи о покрытии;
- после выбора ресурса задание поступает в пакет заданий, который формируется для ресурса R_i, и находится в нем до тех пор, пока он не будет полностью сформирован. По окончании его формирования все задания пакета поступают на выполнение на соответствующий ресурс;
- после того, как задания на ресурсе выполнятся, пул опять полностью заполняется заданиями и т.д. до тех пор, пока все задания из глобальной очереди не будут выбраны, размещены на ресурсы и решены.

Таким образом, новизна предлагаемого подхода заключается:

- в использовании процесса планирования распределения ресурсов на основе предлагаемого метода планирования (в отличие от работы *GRID-систем*, работающих непосредственно с очередями, в которых реализован централизованный планировщик (например, FCFS, SJF). Обсуждения преимуществ такого подхода изложены в работах [3, 4, 6, 8]. Учитывая, что начало следующего действия обработки каждого задания непосредственно начинается после окончания предыдущего и происходит только в заданной последовательности, можно сделать вывод о том, что в условиях нестационарности поступающих заданий предлагаемая модель относится к классу дискретно-событийных имитационных моделей. Отметим, что такие модели рассмотрены в работах [6, 8] и показали свою высокую эффективность при использовании в *GRID*-системах.

Отметим, что пакетный режим обработки на основе буфера достаточно широко используется в устройствах телекоммуникационных и компьютерных систем для повышения эффективности планирования их работы (коммутаторы, маршрутизаторы, сетевые адаптеры). Преимущества использования буфера проявляется в условиях высокой интенсивности данных, поступающих на устройства. Вопросы применения буферов в *GRID* и компьютерных сетях исследовались в работах [6 – 8].

Введем следующие определения, позволяющие уточнить и упростить понимание сущности моделируемых процессов.

Под жизненным циклом задания будем понимать последовательность этапов его обработки – начиная с момента времени помещения его в глобальную очередь, и заканчивая временем его решения на выбранном ресурсе.

Под жизненным циклом глобальной очереди заданий будем понимать совокупность жизненных циклов всех заданий глобальной очереди, для него необходимо определить все характеристики производительности работы модели *GRID*. Отметим, что введение этих понятий обусловлено тем, что анализ производительности дискретно-событийной модели системы необходимо осуществить на разрешении заданий глобальной очереди – сформированной и обслуженной *за определенный промежуток времени работы системы*.

Будем рассматривать гетерогенную *GRID*-систему как среду, состоящую из различных типов ресурсов $\{R_i\}$, и в которую поступает поток различных типов заданий $\{З_i\}$ (рис. 1).

Рис. 1. *GRID* как гетерогенная среда

В общей очереди заданий каждое задание характеризуется вектором характеристик $(\alpha_1^3, \alpha_2^3, ..., \alpha_m^3)$, в качестве которых могут быть тип процессора, объем требуемой для его выполнения памяти, требования к времени решения и т.д. Гетерогенность системы проявляется в том, что одно задание из множества $\{З_i\}$ может выполниться на нескольких из множества $\{R_i\}$ ресурсах, что позволяет использовать это свойство системы для формализации предлагаемого в данной работе подхода. Распределение заданий осуществляется на основе принципа раздельного распределения заданий на ресурсы (рис. 2).

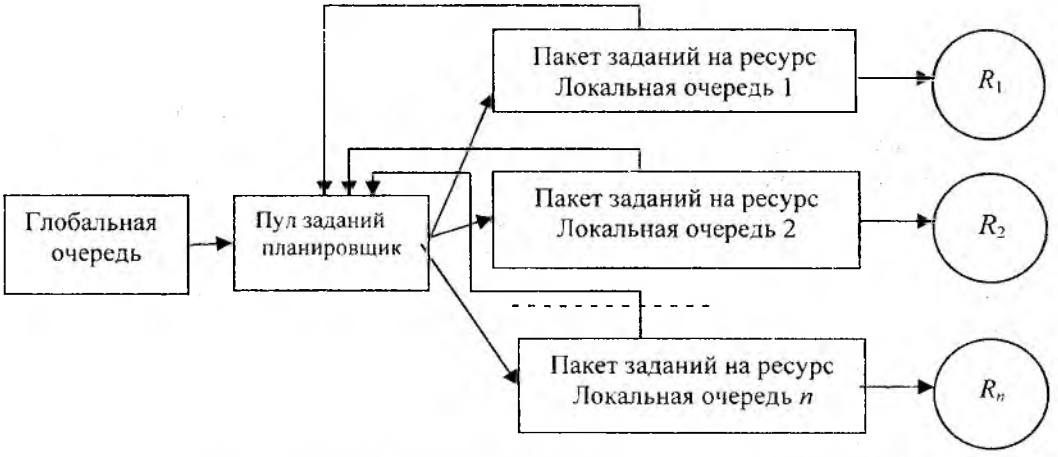


Рис. 2. Раздельное распределение заданий в гетерогенной среде

В этих условиях распространенный метод типа *FCFS* и его модификации приводят к существенной неравномерности загрузки ресурсов [6, 8].

Рассмотрим процесс формирования пакетов – очередей на локальные ресурсы – из глобальной очереди, имеющих ограниченные размеры. Формирование очередей возможно реализовать с помощью таблицы соответствия, показывающей, какие задания на каком ресурсе могут быть решены (выполнены) с учетом того, что на одном ресурсе может выполняться несколько независимых заданий (см. табл. 1, 2).

В ячейках $\{i,j\}$ таблицы соответствия проставляется значение «1» в случае, если задание $З_i$ может быть выполнено на ресурсе R_i В случае использования для планирования алгоритмов типа *FCFS* очередь формируется последовательно, задания отправляются в соответствии с их приоритетом на имеющийся на данный момент времени свободный ресурс непосредственно после нахождения первого на данный момент свободного ресурса из списка доступных ресурсов. *Свободным ресурсом* для данного подхода является такой, для которого в локальной очереди на данный ресурс есть свободное место в пакете, т.е. пакет для этого ресурса не сформирован полностью. При этом сначала в очередь помещаются все задания, которые могут быть решены на первом ресурсе, далее случайным образом выбирается следующий ресурс и аналогично формируется следующая локальная очередь на ресурс и т.д. до тех пор, пока пакеты для формирования локальных очередей на ресурсы будут заполнены (сформированы).

Для упорядочивания формирования локальных очередей положим, что сначала формируется первая локальная очередь (очередь на первый ресурс), далее вторая и т.д. пока все очереди на ресурсы не будут сформированы. Необходимо выбрать некоторую стратегию формирования очередей, обеспечивающую эффективное использование ресурсов системы. Особенностью работы гетерогенной системы с большим количеством ресурсов является то, что некоторые задания могут быть выполнены только на определенных типах ресурсов, а другие – на этих же типах ресурсов и еще дополнительно на нескольких типах ресурсов. Поэтому при произвольном размещении заданий в очереди по таблице соответствия может возникать конкуренция некоторых заданий к одному ресурсу. Это может привести к неравномерности загрузки ресурсов и существенной задержке выполнения конкурирующих заданий. Для устранения этих проблем предлагается распределять задания в очереди на предыдущем шаге таким образом, чтобы подготовить на последующем шаге максимально большее количество свободных ресурсов для выполнения на них следующего подмножества заданий из пула.

Для этого предлагается метод отображения процесса решения заданий глобальной очереди на локальные очереди на ресурсы на основе непрерывного решения задачи о наименьшем покрытии (ЗНП).

Таблица 1

Соответствия, определяющая какие задания на каких ресурсах могут быть решены

	R_1*	R_2	R_3*	R_4	R_5*	R_6	R_7	R_8
$З_1$	1							
$З_2$		1	1	1				
$З_3$	1					1		
$З_4$			1					1
$З_5$		1	1	1				
$З_6$	1						1	
$З_7$					1			1
$З_8$					1		1	
$З_9$					1			
$З_{10}$		1	1					1
$З_{11}$		1			1			
$З_{12}$	1			1		1		

Таблица 2

Соответствия, определяющая какие задания на каких ресурсах могут быть решены

	R_1*	R_2*	R_3	R_4*	R_5	R_6	R_7*	R_8
$З_{13}$	1							
$З_{14}$		1						
$З_{15}$					1		1	1
$З_{16}$				1				
$З_{17}$			1				1	1
$З_{18}$		1						
$З_{19}$		1						
$З_{20}$			1			1	1	
$З_{21}$					1	1	1	1
$З_{22}$	1							
$З_{23}$			1	1			1	1
$З_{24}$		1						

Модель планирования на основе задачи о наименьшем покрытии, вычислительный эксперимент и анализ полученных результатов

Случайное формирование локальных очередей на основе метода *FCFS* в гетерогенных структурах может привести к большой неравномерности загрузки ресурсов и их неэффективному использованию.

Распределение заданий рассмотрим на основе принципа раздельного распределения заданий (см. рис.2), но при этом распределение с помощью диспетчера будет осуществляться не статично, как это делается в известных раздельных схемах распределения заданий, а динамично и непрерывно на основе следующей процедуры *D* .

Процедура D

1.На основе множества заданий пула, информации о заданиях и свободных ресурсах планировщиком первого уровня формируется таблица соответствия.

2. На основе таблицы соответствия, решения задачи о наименьшем покрытии (ЗНП) определяется *минимальное количество ресурсов*, на котором очередь заданий, вошедших в пул, может быть выполнена. При этом сначала заполняется очередь на ресурс, который вошел в минимальное покрытие и на нем возможно решение наибольшего количества заданий, далее очередь формируется к ресурсу из покрытия с наибольшими возможностями по решению задач среди оставшихся ресурсов, принадлежащих покрытию и т.д. Если очередь начинает превышать размер пакета заданий на ресурс (т.е. пакет заданий на данный ресурс

сформирован), задания в этом случае помещаются на свободный ресурс, который может его выполнить. Если таких ресурсов уже нет, то задание отправляется обратно в пул заданий.

3. Далее пул заданий заполняется новыми заданиями из глобальной очереди, и процедура повторяется. Благодаря применению в динамической процедуре *D* при планировании очередей ЗНП обеспечивается равномерная загрузка каждого ресурса. Далее реализация планирования выполнения заданий в локальных очередях к ресурсам может быть осуществлено программными средствами локальных планировщиков (PBS).

Для практической реализации процедуры *D* был выбран эффективный приближенный алгоритм решения ЗНП, рассмотренный в работе [6]. Для проверки эффективности такого подхода к планированию распределения ресурсов была разработана программная реализация разработанной модели *GRID_Scheduler_Model*, реализующая имитационную дискретно-событийную модель функционирования гетерогенной среды (см. рис. 3), включающую следующие функциональные элементы:

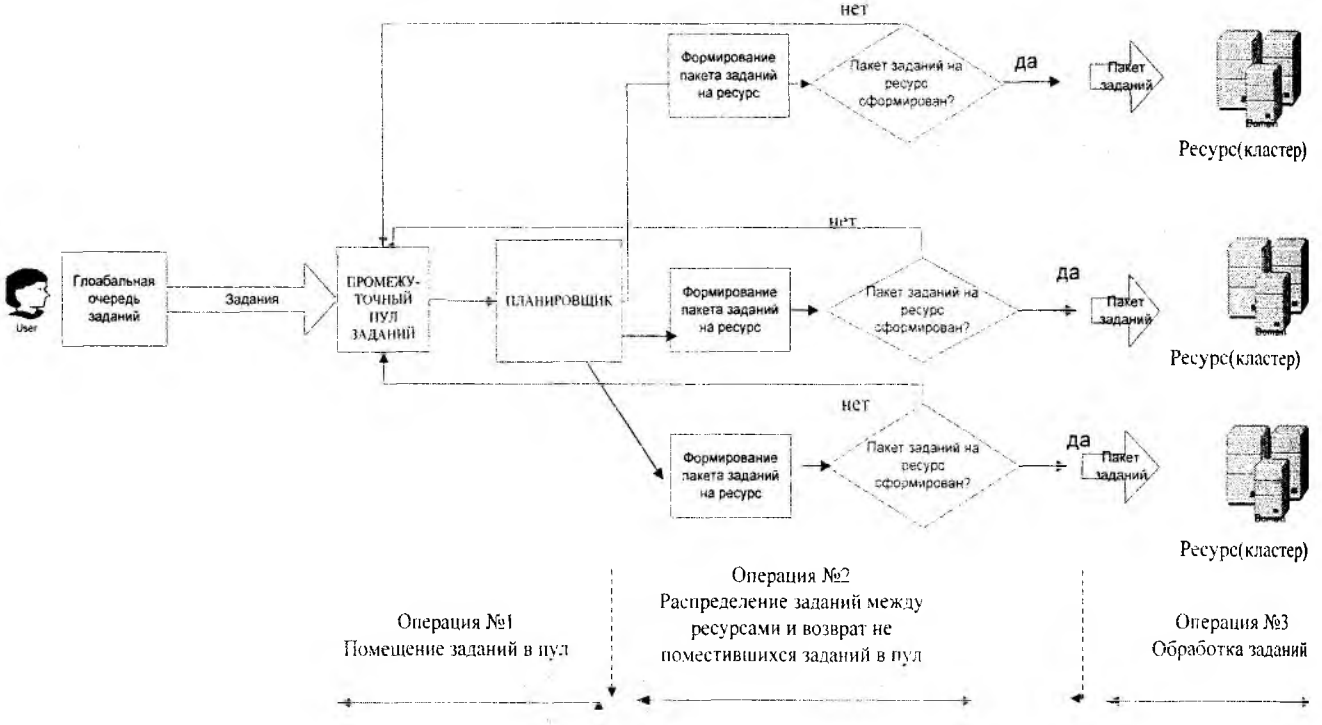


Рис. 3. Общая схема работы модели

Глобальная очередь заданий – задания (заявки, требования, приложения), которые подаются на систему для моделирования процесса работы гетерогенной среды.

Промежуточный пул заданий – буфер системы, в который в порядке поступления загружаются задания из глобальной очереди и *осуществляется планирование ресурсов*.

Планировщик – программный модуль, который распределяет на основе выбранного метода планирования задания, находящиеся в пуле, на ресурсы.

Модель заданий – определяется *сложностью* заданий, задается в тактах – более сложное задание требует для его выполнения большего количества тактов.

Ресурсы – элементы, моделирующие процесс выполнения заданий, которые были на них распределены, *производительность* которых тоже определяется числом тактов необходимых для освобождения ресурсов.

Модель времени – в качестве модели времени была выбрана условная единица времени работы системы – такт.

Один такт включает выполнение системой трех операций:

Операции № 1 – помещение входных заданий в пул;

Операция № 2 – распределение задач в очереди пула между ресурсами и возврат непоместившихся на ресурсы заданий обратно в пул;

Операция № 3 – имитация решения заданий ресурсами.

Для моделирования планирования использовалась матрица в виде прямоугольной таблицы, строки которой соответствуют номерам заданий, столбцы – номерам ресурсов, при этом, если значение ячейки таблицы $m_{ij} = 1$, то это означает, что i-е задание решается j-м ресурсом. Количество ресурсов, которыми может быть решено m_i-е задание, задается по одному из случайных законов (равномерному, Пуассона, Эрланга). Максимальное количество ресурсов, которым может быть решено m_i-е задание, определяется параметром «*универсальность задачи*», который задается как определенный процент общего количества ресурсов, на которых может быть выполнено задание. Модель также позволяет оценивать количество заданий, попадающих на каждый ресурс системы, время нахождения заданий в системе, а также оценивать временные выигрыши по времени выполнения всех заданий в очереди при использовании различных процедур планирования и варьировать такими параметрами как сложность решаемых задач, производительность отдельных ресурсов, задержка заданий при их отправке на нижний уровень, частотой планирования выбранной процедурой планирования.

В программе *GRID_Scheduler_Model* реализованы следующие методы планирования:

FCFS (First Come First Served) – метод распределяет задания по принципу «первый пришел первым обслужился»(*FIFO*), то есть в порядке их поступления (в порядке очереди), выбирая при этом для задания свободный и доступный ресурс, и централизованный метод планирования ресурсов в *GRID* [20]. Планировщик выбирает первую задачу из очереди пула и пробует поместить ее в пакет заданий на один из ресурсов, на которых она может быть решена. Если операция выполнена успешно, то планировщик переходит к новому заданию, если нет – возвращает его в пул, при этом оно остается первым в пуле и переходит к следующему заданию из очереди.

MC (Minimal Cover) – метод основан на поиске наименьшего количества столбцов, которыми можно покрыть все строки в матрице (ЗНП). Задания, находящиеся в пуле, отображаются в виде прямоугольной матрицы соответствий, в которой строки определяют номера заданий, а столбцы – номера ресурсов. Если i-е задание может быть решено на j-м ресурсе, то в ячейке ij матрицы стоит значение «1» (см. табл. 1, 2). Метод находит наименьшее количество столбцов (ресурсов), которые могут покрыть (решить) все строки (задания), находящиеся в пуле. После этого задания распределяются на ресурсы, вошедшие в минимальное покрытие, и решаются на ресурсах. Если пакеты заданий на соответствующие ресурсы сформированы, а нераспределенные задания еще остались, то распределяем их на другие свободные ресурсы, а оставшиеся нераспределенными на ресурсы задания возвращаем обратно в пул и опять решаем задачу о покрытии.

Постановка эксперимента и метрики производительности работы системы

В качестве метрик производительности работы модели использовались коэффициент использования ресурсов и время ответа системы (время выполнения всех заданий глобальной очереди); коэффициент выигрыша по времени выполнения всех заданий в очереди:

$$K_a = \frac{T_{\sum}\ (FIFO)}{T_{\sum}\ (MC)},$$

где $T_{\sum}\ (FIFO)$ – время выполнения очереди заданий при использовании в качестве планировщика процедуры *FIFO*; $T_{\sum}\ (MC)$ – время выполнения той же очереди заданий при использовании в качестве планировщика процедуры *MC*. При моделировании исследовались зависимости времени выполнения всех заданий глобальной очереди в зависимости от числа

заданий в очереди при различных сложностях решаемых заданий и различной производительности ресурсов, а также зависимость коэффициента использования ресурсов гетерогенной *GRID*-системы от количества решаемых заданий. При этом коэффициент использования гетерогенной *GRID*-системы рассчитывался как отношение суммы количества тактов, выполненных ресурсами, работающими на каждом такте выполнения заданий, к общему количеству тактов, которое может быть выполнено всеми ресурсами системы за время выполнения всех заданий глобальной очереди. Анализ полученных зависимостей проведен для различных законов поступления и обслуживания заданий – равномерного закона, законов Пуассона и Эрланга.

Планировщик на основе процедуры *MC* планирует существенно лучше, чем планировщик на основе процедуры *FIFO*, что выражается в том, что он позволяет более равномерно загружать ресурсы системы.

Его применение позволяет уменьшить время выполнения всех заданий. На рис.6 приведена зависимость коэффициента использования 10 гетерогенных кластеров, имеющих по десять однородных ресурсов, при этом

$$K_a = \frac{T_\Sigma\ (FIFO)}{T_\Sigma\ (MC)} = \frac{65678}{20091} = 3{,}23\ ,$$

где $T_\Sigma\ (FIFO)$ – время выполнения очереди заданий при использовании в качестве планировщика процедуры *FIFO*; $T_\Sigma\ (MC)$ – время выполнения той же очереди заданий при использовании в качестве планировщика процедуры *MC* заданы в тактах работы ресурсов. Экспериментальное исследование показало, что данный выигрыш может быть равен 10 и более. С увеличением числа ресурсов неравномерность их загрузки на основе процедуры планирования типа *FIFO* резко возрастает, в случае применения процедуры MC практически не изменяется.

Из исследования влияния различных законов распределения изменения параметров системы на характеристики планирования следует, что наиболее «тяжелым», с точки зрения планирования выполнения заданий, является равномерный закон распределения параметров системы.

Поэтому для этого закона распределения приведены результаты исследования взаимосвязи таких характеристик планирования, как время разрешения очереди, коэффициент использования ресурсов системы, размер очереди и размер буфера.

Сравнительный анализ планировщиков на основе процедуры *MC* и *FIFO* при их работе в гетерогенных структурах показал, что, подбирая размер буфера можно обеспечить выигрыш во времени работы выполнения заданий планировщиком на основе *MC* и при этом он обеспечивает равномерность загрузки всех ресурсов, а планировщик на основе процедуры планирования *FIFO* оставляет незагруженными от 30 до 40 % ресурсов.

Заключение

В итоге можно сделать важные для последующих исследований выводы:

-аналитическое выражение поверхности зависимостей имеет квадратичный характер, что можно использовать при одновременном управлении и планировании системы;

-функции поверхностей – выпуклы и имеют локальный минимум, следовательно, можно определить оптимальные соотношения между моделируемыми параметрами на основе получения значений, доставляющих минимум приведенным функциям, решая систему уравнений для частных производных по отдельным параметрам;

-задавая требуемые ограничения на значения коэффициента загрузки и среднее время ответа системы, можно моделировать и прогнозировать значения отдельных параметров

моделируемой системы, например размеры буфера и пула, в зависимости от интенсивности входного потока заданий, поступающего в глобальную очередь *GRID*.

Таким образом, в работе рассмотрен подход к планированию распределением ресурсов вычислительных ресурсов в двухуровневой гетерогенной *GRID*-системе на основе решения ЗНП. Предложена базовая модель и общая схема работы двухуровневой гетерогенной *GRID*-системы, использующая таблицу соответствия назначения заданий на выбранные в результате планирования ресурсы, и показано, что для нее можно использовать подход к планированию выбора ресурсов на основе ЗНП. Основные преимущества рассмотренного подхода заключаются в реализации процесса планирования на основе новых функциональных элементов – пула заданий, выбираемых из глобальной очереди, и пакета заданий на ресурсы, в совокупности позволяющих повысить равномерность загрузки ресурсов и оптимизировать время выполнения всей очереди заданий. Вычислительные эксперименты, проведенные с использованием программной реализации модели, подтвердили эффективность предложенного подхода по сравнению с известным и используемым в современных ***GRID***-системах методом *FIFO* с учетом выбранных метрик производительности системы.

Список литературы: 1. *Загородній, А.Г.* Грід – нова інформаційна обчислювальна технологія для науки / А.Г. Загородній, Г.М. Зіновєв, Є.С. Мартинов, С.Я. Свістунов, В.М. Шадура // Вісник НАН України. – 2005. – №6. – С. 17–25., 2. *Zagorodny, A., Zgurovsky, M., Zinoviey, G., Petrenko, A., Martynov, E.* Integrated Ukraine into European Grid infrastructure // Системні дослідження і інформаційні технології. – 2009. – № 2. – С. 35–49. 3. *Мирошник, М.А.* Решение задач диспетчеризации в распределенных телекоммуникационных системах / М.А. Мирошник, В.Г. Котух, С.Н. Селевко // Радиотехника. – 2011. – Вып. 169. – С. 139–152., 4. *Мирошник, М.А.* Отказоустойчивость распределенных телекоммуникационных систем / М.А. Мирошник, В.Г. Котух, С.Н. Селевко // Радиотехника. – 2011. – Вып. 168. – С. 51–55. 5. *Шелестов, А.Ю.* GRID-технологии в системах мониторинга на основе спутниковых данных / А.Ю. Шелестов, Н.Н. Куссуль, С.В. Скакун // Проблемы управления и информатики. – 2006. – №1,2. – С. 259 – 270. 6. *Мирошник М.А.* Методы повышения отказоустойчивости телекоммуникационных систем / В.Г. Котух, М.А. Мирошник, С.Н.Селевко // Радиотехника. – 2011. – Вып. 166. – С. 259–268. 7. *Коваленко, В. Н.* Методы и алгоритмы управления параллельными заданиями в гриде с ресурсами в форме кластеров / В. Н. Коваленко, Д. А. Семячкин // Вестник Южного научного центра РАН. – 2008. – Т. 4. – № 3. – С. 23–34. 8. *Мирошник, М.А.* Развитие современных направлений цифровых телекоммуникационных систем и сетей / В.Г. Котух, М.А. Мирошник, Селевко С.Н. // Радиотехника. – 2011. – Вып. 165. – С. 254–258. 9. *Коваленко, В.Н.* Грид-диспетчер: реализация службы диспетчеризации заданий в Грид / В.Н. Коваленко, Е.И. Коваленко, Д.А. Корягин, Э.З. Любимский, Е.В. Хухлаев, О.Н. Шорин // Сб. докл. Междунар. конф. «Распределенные вычисления и Грид-технологии в науке и образовании», Дубна, 2004. 10. *Минухин, С.В.* Моделирование планирования ресурсов GRID средствами пакета GridSim / С.В. Минухин, А.В. Коровин // Системи обробки інформації. Інформаційні технології та комп'ютерна інженерія. – 2011. – Вип. 3(93). – С. 62 – 68.

Харьковская национальная академия
городского хозяйства
Харьковская государственная академия
железнодорожного транспорта
Харьковский национальный
университет радиоэлектроники

Поступила в редколлегию 21.09.2012

УДК 621.372.852

Л.М. КАРПУКОВ, д-р техн. наук, Р.Ю. КОРОЛЬКОВ

ПРЯМОЙ МЕТОД СИНТЕЗА ПОЛОСНО-ПРОПУСКАЮЩИХ ШЛЕЙФОВЫХ ФИЛЬТРОВ С ЧЕБЫШЕВСКОЙ ХАРАКТЕРИСТИКОЙ

Введение

Широкое использование в радиотехнических и телекоммуникационных системах микроволновых фильтров обуславливает потребность в совершенствовании существующих и в разработке новых методов их синтеза. Существующие методы синтеза микроволновых фильтров основаны на применении фильтров-прототипов. Наибольшее распространение в процедурах синтеза в качестве прототипа получил фильтр с лестничной структурой на элементах с сосредоточенными постоянными. Переход от фильтра-прототипа к структуре микроволнового фильтра, реализуемой на элементах с распределенными параметрами, в этих процедурах осуществляется с помощью частотного преобразования Ричардса и операций по введению в структуру фильтра-прототипа единичных элементов в виде четвертьволновых отрезков передающих линий [1 – 3]. В [4] в качестве прототипа использован ступенчатый переход с последующей заменой, как правило, приближенной, звеньев перехода звеньями иной конфигурации, из которых составляется структура синтезируемого фильтра. В [5] применен цифровой фильтр как фильтр-прототип. Переход от цифрового фильтра к микроволновому фильтру осуществляется путем решения оптимизационной задачи.

В настоящей работе предложен метод прямого синтеза шлейфовых полоснопропускающих фильтров с чебышевской характеристикой передачи, не требующий использование фильтров-прототипов. Предполагается, что структура фильтра симметрична и составлена из четвертьволновых отрезков линий передачи и короткозамкнутых шлейфов. Предлагаемый метод обеспечивает составление непосредственно по техническому заданию равноволновой функции фильтрации, определяющей структуру и соотношения для параметров рассеяния фильтра. Для нахождения волновых сопротивлений шлейфов и отрезков линий синтезируемого фильтра предложена простая процедура, основанная на перемножении матриц.

Функции фильтрации

Структуры исследуемых фильтров представлены на рис. 1, где ρ_0, ρ_L, ρ_S – волновые сопротивление нагрузок, отрезков линий, короткозамкнутых шлейфов, соответственно.

Математические модели элементов фильтров определим следующим образом:

$$\left.\begin{aligned} M_L(\theta)=\begin{bmatrix} A_L(\theta) & -A_{L11}(\theta) \\ A_{L11}(\theta) & -B_L(\theta) \end{bmatrix} &= -\cos(\theta)\cdot\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}+j\cdot\frac{\sin(\theta)}{2R_L}\cdot\begin{bmatrix} \rho_L{}^2+1 & -1\cdot(\rho_L{}^2-1) \\ \rho_L{}^2-1 & -1\cdot(\rho_L{}^2+1) \end{bmatrix}, \\ A_{L21}&=1, \end{aligned}\right\} \tag{1}$$

- для отрезка линии;

$$\left.\begin{aligned} M_S(\theta)=\begin{bmatrix} A_S(\theta) & -A_{S11}(\theta) \\ A_{S11}(\theta) & -B_S(\theta) \end{bmatrix} &= -\sin(\theta)\cdot\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}+j\cdot\frac{\cos(\theta)}{2R_S}\cdot\begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix}, \\ A_{S21}(\theta)&=\sin(\theta), \end{aligned}\right\} \tag{2}$$

- для короткозамкнутого шлейфа.

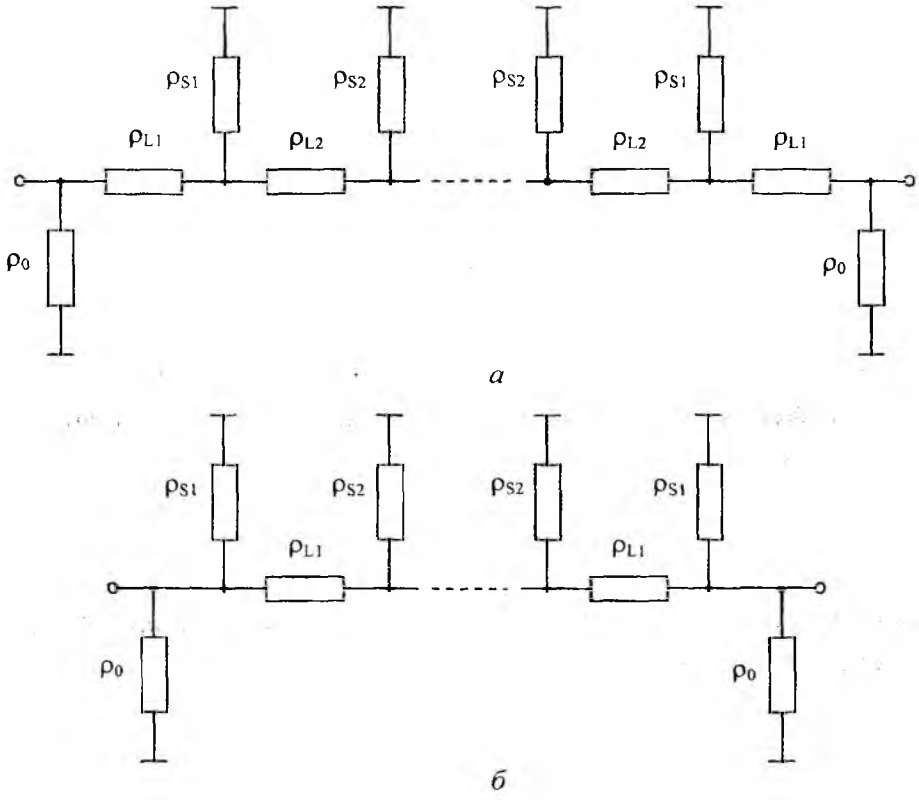


Рис. 1. Структуры фильтров

Здесь $j=\sqrt{-1}$ – мнимая единица, θ – электрическая длина отрезков линий; A_{ik} - числитель, B – знаменатель соответствующего параметра рассеяния $S_{ik}=A_{ik}/B$; $A=\left(A_{11}^2-A_{21}^2\right)/B$; $R_L=\rho_L/\rho_0$, $R_S=\rho_S/\rho_0$ – нормированные волновые сопротивления.

Характеристики фильтров вычисляются путем перемножения характеристик его элементов [6]:

$$\left.\begin{aligned}\hat{M}(\theta)&=\prod_{k=1}^{n}M(\theta)_k,\\ \hat{A}_{21}(\theta)&=(-1)^{n-1}\prod_{k=1}^{n}A_{21_k}(\theta).\end{aligned}\right\}, \tag{3}$$

Структура матрицы $\hat{M}(\theta)$ аналогична структуре матриц M_S, M_L.

Для фильтров без потерь имеет место:

$$\left|\hat{S}_{21}(\theta)\right|^2=\frac{1}{1+\left|F(\theta)\right|^2},$$

Здесь $F(\theta)=\hat{A}_{11}(\theta)/\hat{A}_{21}(\theta)$ – функция фильтрации, представляющая собой отношение числителей параметров рассеяния фильтра $\hat{S}_{11}=\hat{A}_{11}/\hat{B}$, $\hat{S}_{21}=\hat{A}_{21}/\hat{B}$. В общем случае функция фильтрации комплексная, для устройств, обладающих симметрией, она становится мнимой функцией.

Анализ структур фильтров на рис. 1 при различном количестве элементов по соотношениям (1) – (3) позволил получить следующие соотношения:

$$\hat{A}_{21}(\theta)=\sin(\theta)^{n_S}, \tag{4}$$

$$\hat{A}_{11}(\theta) = j\sin(\theta)^{n_S-1}\hat{a}_{11}(\theta), \tag{5}$$

где

$$\hat{a}_{11}(\theta) = \begin{cases} \cos(\theta)\sum\limits_{k=0}^{n_L/2}\alpha_k\cos(\theta)^{2k}, n_L - \textit{четное}, \\ \sum\limits_{k=0}^{(n_L+1)/2}\alpha_k\cos(\theta)^{2k}, n_L - \textit{нечетное}. \end{cases} \tag{6}$$

Здесь n_L – количество отрезков линий, n_S – количество шлейфов.

Таким образом, функция фильтрации рассматриваемых структур фильтров запишется, без учета мнимой единицы, в виде

$$F(\theta) = \frac{\hat{a}_{11}(\theta)}{\sin(\theta)}. \tag{7}$$

Полученные соотношения (6), (7) с точностью до постоянного коэффициента соответствуют Чебышевской функции вида

$$Ch(\theta, n) = \cos\left\{\arccos\left[\frac{ctg(\theta)}{ctg(\theta_c)}\right] + n\cdot\arccos\left[\frac{\cos(\theta)}{\cos(\theta_c)}\right]\right\}, \tag{8}$$

где $n = n_L/2$ – для четного числа отрезков или $n_L + 1$ – для нечетного, θ_c – угол, соответствующий границе полосы пропускания по уровню пульсаций фильтра с равноволновой амплитудно-частотной характеристикой.

Для младших значений порядка n эти функции записываются в полиномиальной форме следующим образом:

$$Ch(\theta,1) = \left|\left(1+\sin(\theta_c)\right)c(\theta)^2 - 1\right| / \sin(\theta), \tag{9а}$$

$$Ch(\theta,2) = c(\theta)\left|2\left(1+\sin(\theta_c)\right)c(\theta)^2 - \left(2+\sin(\theta_c)\right)\right| / \sin(\theta), \tag{9б}$$

$$Ch(\theta,3) = \left|4\left(1+\sin(\theta_c)\right)c(\theta)^4 - \left(5+3\sin(\theta_c)\right)c(\theta)^2 + 1\right| / \sin(\theta), \tag{9в}$$

$$Ch(\theta,4) = c(\theta)\left|8\left(1+\sin(\theta_c)\right)c(\theta)^4 - 4\left(3+2\sin(\theta_c)\right)c(\theta)^2 + \left(4+\sin(\theta_c)\right)\right| / \sin(\theta), \tag{9г}$$

$$Ch(\theta,5) = \left|16\left(1+\sin(\theta_c)\right)c(\theta)^6 - 4\left(7+5\sin(\theta_c)\right)c(\theta)^4 + \left(13+5\sin(\theta_c)\right)c(\theta)^2 - 1\right| / \sin(\theta), \tag{9д}$$

где $c(\theta) = \cos(\theta)/\cos(\theta c)$.

Методика синтеза

Процедура синтеза состоит из двух этапов. На первом этапе из числителей и знаменателя S-параметров фильтра составляется матрица, обеспечивающая заданные характеристики, на втором этапе определяются волновые сопротивления элементов фильтра.

Первый этап включает в себя нахождение порядка n аппроксимирующей функции (8) по заданным затуханиям α_c в полосе пропускания и $\alpha_з$ при $\theta_з$ в полосе заграждения, выраженным в децибеллах:

$$n = \frac{arcch\left(\sqrt{\frac{10^{\frac{\alpha_з}{10}}-1}{10^{\frac{\alpha_c}{10}}-1}}\right) - arcch\left(\frac{ctg(\theta з)}{ctg(\theta c)}\right)}{arcch\left(\frac{\cos(\theta з)}{\cos(\theta c)}\right)}. \quad (10)$$

По требуемой полосе пропускания $\Delta\theta$ вычисляется угол, соответствующий границе полосы пропускания:

$$\theta_c = \frac{\pi}{2}\left(1 - \frac{\Delta\theta}{2}\right). \quad (11)$$

Составляется функция фильтрации:

$$F(\theta) = \frac{\varepsilon Ch(\theta, n)}{\sin(\theta)}, \quad (12)$$

где $\varepsilon = \sqrt{10^{\frac{\alpha_c}{10}} - 1}$.

Выбирается структура фильтра и по (4), (5) определяются числители S-параметров фильтра. Составляется квадрат модуля знаменателя S-параметров:

$$\left|\hat{B}(\theta)\right|^2 = \sin(\theta)^{2(n_S-1)}\varepsilon^2 Ch(\theta, n)^2 + \sin(\theta)^{2n_S}. \quad (13)$$

Вычисляются корни этой функции с последующей ее записью в виде произведения множителей:

$$\left|\hat{B}(\theta)\right|^2 = \gamma \prod_{k=0}^{n_L+n_S}\left[\cos(\theta)^2 - \cos(\theta_k)^2\right]. \quad (14)$$

Осуществляется переход в (14) к переменной $z = \exp(-j2\theta)$. В функции $|B(z)|^2$ выделяются корни $z_k = \exp(-j2\theta_k)$, удовлетворяющие условию $|z_k| \geq 1$, и составляется знаменатель S-параметров:

$$\hat{B}(z) = k_0 \prod_{k=0}^{n_L+n_S}[z - z_k]. \quad (15)$$

Оставшиеся корни, удовлетворяющие условию $|z_k| \leq 1$, будут принадлежать функции, комплексно сопряженной с $B(z)$. Заключительной операцией первого этапа является формирования матрицы синтезируемого фильтра:

$$\hat{M}(z) = \begin{bmatrix} \hat{A}(z) & -\hat{A}_{11}(z) \\ \hat{A}_{11}(z) & -\hat{B}(z) \end{bmatrix} \quad (16)$$

Определение волновых сопротивлений элементов фильтра на втором этапе процедуры синтеза производится по значению коэффициента отражения $\hat{S}_{11}(z)$ при $z = 0$ для исходного соединения, соответствующего структуре фильтра, и всех промежуточных соединений,

возникающих при удалении элементов фильтра. Если крайний элемент соединения – отрезок линии, то его нормированное волновое сопротивление вычисляется по формуле

$$R_L = \frac{1+\hat{S}_{11}(0)}{1-\hat{S}_{11}(0)}. \qquad (17)$$

В случае, когда по краям находится шлейф, расчет ведется на основании формулы

$$R_{экв} = \frac{1+\hat{S}_{11}(0)}{1-\hat{S}_{11}(0)} = \frac{R_S R_L}{R_S + R_L}. \qquad (18)$$

По этой формуле могут быть получены разные комбинации из значений волновых сопротивлений шлейфа и соединенного с ним отрезка линии.

После нахождения волновых сопротивлений крайних элементов, они удаляются, что сопровождается умножением матрицы соединения на матрицы, обратные матрицам удаляемых элементов.

Пример синтеза фильтра

Выполним синтез фильтра по исходным данным из [3]: $\alpha_c = 0.1$ дб; $\Delta\theta = 0.5$; число шлейфов $n_S = 5$, число отрезков линий $n_L = 4$, структура фильтра соответствует рис. 1,б.

Для указанных данных функция фильтрации (12) будет представлена функцией (9г) при $\varepsilon = 0.1526$, $\theta_c = 1.1781$. Числители S-параметров фильтра приобретут следующий вид: $\hat{A}_{21}(\theta) = \sin(\theta)^5$, $\hat{A}_{11}(\theta) = j\sin(\theta)^4 Ch(\theta,4)$, а коэффициенты знаменателя $\hat{B}(z)$, составленного по (13) – (15), получат следующие значения: $k_0 = 0.9009$, $z_0 = -1.51826$, $z_{1,2} = -1.22407 \pm j0.69802$, $z_{3,4} = -0.75537 \pm j0.86861$, $z_{5,6,7,8} = 1$. Этим значениям соответствуют следующие значения нормированных волновых сопротивлений элементов фильтра: $R_{S1} = 0.548$, $R_{L1} = 0.8302$, $R_{S2} = 0.345$, $R_{L2} = 0.8211$, $R_{S3} = 0.3466$.

Аналогичной функцией фильтрации обладает фильтр со структурой на рис. 1,*а* с числом шлейфов $n_S = 3$ и числом отрезков линий $n_L = 4$. В процессе синтеза для этой структуры получены следующие значения нормированных волновых сопротивлений элементов фильтра: $R_{L1} = 0.330$, $R_{S1} = 0.045$, $R_{L2} = 0.119$, $R_{S2} = 0.043$.

В [3] приведены результаты синтеза рассматриваемого фильтра по традиционной методике с помощью фильтра-прототипа: $R_{S1} = 0.567$, $R_{L1} = 0.7731$, $R_{S2} = 0.2883$, $R_{L2} = 0.7176$, $R_{S3} = 0.2931$.

На рис. 2 указаны результаты расчета зависимости коэффициента передачи $\hat{S}_{21}(\theta)$ фильтра от $\theta = 2\pi f/4f_0$, f_0 –центральная частота полосы пропускания. Кривая 1 соответствует расчетам для двух вариантов структур фильтра по предложенному методу, кривая 2 – по традиционной методике.

Как следует из рис. 2 синтез по предлагаемому методу обеспечивает строго равноволновую характеристику. Характеристика, получаемая по традиционной методике синтеза, отличается от равноволновой, причем отличие увеличивается при увеличении полосы пропускания. Если для изготовления фильтров использовать микрополосковую технологию, позволяющую реализовывать микрополосковые линии с волновым сопротивлением в диапазоне от 15 до 116 *Ом* [5], то при пятидесятиомном волновом тракте волновые сопротивления, полученные предложенным методом для структуры на рис. 1, *а* оказываются конструктивно невыполнимы. Волновые сопротивления, полученные традиционным методом синтеза, имеют значения на грани реализации. Только набор значений волновых сопротивлений, полу-

ченный предложенным методом для структуры на рис. 1,*б*, оказывается конструктивно реализуемым.

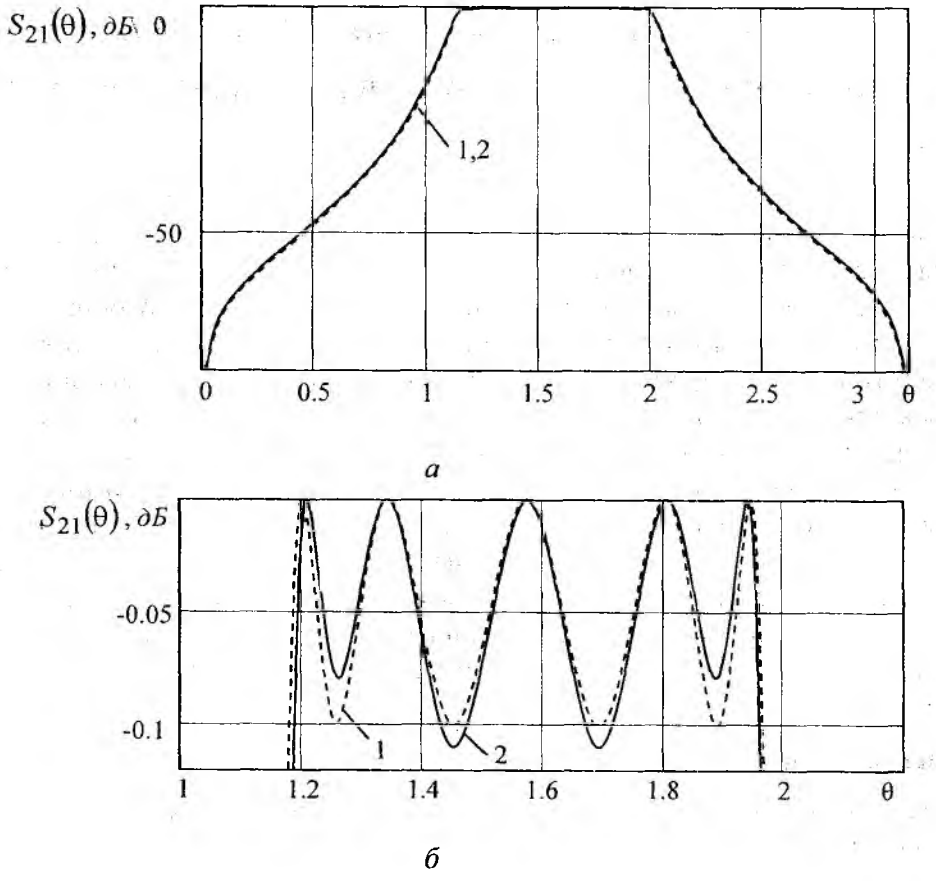


Рис. 2. Амплитудно-частотная характеристика коэффициента передачи фильтра:
1 – расчет фильтра по предложенному методу;
2 – расчет фильтра по традиционной методике

Выводы

Разработан метод прямого синтеза полосно-пропускающих фильтров с равноволновой амплитудной характеристикой, реализуемых на четвертьволновых отрезках линий и короткозамкнутых шлейфах. Метод отличается простотой составления функции фильтрации по заданным техническим требованием, возможностью выбора для полученной функции фильтрации двух вариантов структуры фильтра, а также эффективностью процедуры определения волновых сопротивлений элементов фильтра с возможностью подбора их значений, удовлетворяющих требованиям конструктивной реализации.

Использование предложенного метода синтеза позволит повысить качество проектных работ при разработке селективных устройств микроволнового диапазона.

Список литературы: *Матей, Д.Л., Янг, Л., Джонс, Е. М. Т.* Фильтры СВЧ, согласующие цепи и цепи связи : пер. с анг. ; под ред. Л.В. Алексеева, Ф.В. Кушнира. – М. : Связь, 1971. – Т.1. – 439 с.; 1972. – Т.2. – 493 с. 2. *Современная* теория фильтров и их проектирование / Под ред. Г. Темеша, С. Митры : пер с анг. ; под ред. И.Н. Теплюка. – М. : Мир, 1977. – 560 с. 3. *Hong Jia-Sheng, Lancaster, M.J.* Microstrip filters for RF. Microwave applicatios / Hong Jia-Sheng, Lancaster M.J – New York: John Wiley, 2001. – 476 p. 4.*Фельдштейн, А.Л.* Синтез четырехполюсников и восьмиполюсников на СВЧ / Фельдштейн А.Л., Явич Л.Р. – М.: Связь, 1971. – 388 с. 5. *Lin-Chuan Tsai, Ching-Wen Hsue.* Dual-Band Band-Pass Filters Using Equal-length Coupled-Serial-Shunted Lines and Z-Transform Technique // IEEE Trans. Microwave Theory Tech ., vol. 52, no. 4, 2004 pp.1111-1117. 6. *Карпуков, Л.М.* Символьный анализ устройств СВЧ // Известия вузов. Радиоэлектроника. – 1982.- Т.25.- №6. – С. 85 – 87.

Запорожский национальный технический университет

Поступила в редколлегию 22.09.2012

УДК 621.372(075)

П. Ф. ЛЕБЕДЕВ, канд. техн. наук, В. П. ДРОБЫШЕВА

СХЕМЫ ЗАМЕЩЕНИЯ КОНДЕНСАТОРОВ И КАТУШЕК ИНДУКТИВНОСТИ

Введение

В практике анализа линейных электрических цепей широко используются две двухэлементные схемы замещения. Это так называемые RC и RL-цепи. Детальное рассмотрение структурных вариаций этих цепочек связано с внедрением в расчетную практику компьютерных методов анализа и распространенным мнением, что эти преобразования эквивалентны только в рамках одной частоты и не могут быть использованы в условиях девиации частот. Другая интересная сторона связана с идеей сравнения как однородных, так и разнородных цепочек друг с другом в условиях частотной зависимости при одинаковом подборе параметров на базовой частоте. Кроме того, нам кажется возможным в условиях современного анализа показать различные методики, связанные как с преобразованиями сопротивлений в проводимости, так и наоборот и, особенно, из анализа последствий этих преобразований.

Цель работы – выполнить обстоятельный частотный анализ схем замещения катушек индуктивности и конденсаторов, рассчитанных на базовой частоте и имеющих вполне определённые исходные параметры, подвергающиеся частотно зависимому воздействию. Интерес представляют также некоторые комбинации из рассматриваемых структур в частотном варианте анализа.

Анализ состояния

На рис. 1 представлены все четыре варианта схем замещения RL и RC-цепей в двухэлементном изображении.

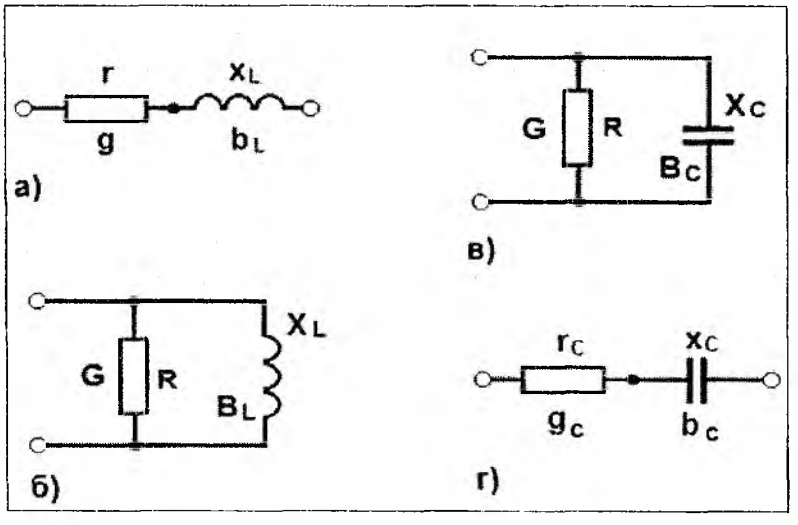


Рис. 1. Схемы замещения катушки индуктивности – слева;
схемы замещения конденсатора – справа

Каждый из элементов схемы обозначен двояко: и как сопротивление и как проводимость. Это сделано с целью устранения необходимости повторного представления схем в разных подходах анализа. С этой же целью обозначения сопротивлений и проводимостей в разных структурах тоже выбраны не одинаково. Реактивные сопротивления и проводимости на схемах также выделены. Но нас интересуют случаи равных соответствующих величин, по этой причине индексы индуктивностей «L» и емкостей «C» иногда будут опущены. При обозначении входных сопротивлений иногда будем придерживаться индексов 1, 2, 3 и 4 в соответствии с порядком обозначения схем буквами: а), б), в), г).

Вариант преобразования с сопротивлениями и проводимостями

Переход от схемы а) к схеме б) или от схемы г) к схеме в) в условиях одной частоты обычно решают просто путём деления соответствующего сопротивления на квадрат модуля входного сопротивления:

$$G = \frac{r}{Z^2};\quad B_L = \frac{x_L}{Z^2};\quad Z^2 = r^2 + x_L^2;\quad \text{или}\quad G = \frac{r}{Z_C^2};\quad B_C = \frac{x_C}{Z_C^2};\quad Z_C^2 = r^2 + x_C^2 \qquad (1)$$

Обратный переход решают аналогично, но сопротивления получают, используя проводимости параллельных схем замещения:

$$r = \frac{G}{Y_L^2};\quad x_L = \frac{B_L}{Y_L^2};\quad Y_L^2 = G^2 + B_L^2;\quad \text{или}\quad r = \frac{G}{Y_C^2};\quad x_C = \frac{B_C}{Y_C^2};\quad Y_C^2 = G^2 + B_C^2; \qquad (2)$$

В некоторых учебных пособиях, например [1], предостерегают, что не следует использовать сопротивления последовательной схемы замещения для определения проводимостей. На наш взгляд, это делать можно, но правильно ими пользоваться. Пусть $g = 1/r$ и $b = 1/x_L$. Тогда в комплексной форме записи получим:

$$\underline{Y} = g \cdot jb/(g + jb) = g \cdot b^2/(g^2 + b^2) - jb \cdot g^2/(g^2 + b^2) = G - jB_L. \qquad (3)$$

Если в уравнениях для проводимостей G и B_L в (1) поменять сопротивления на соответствующие им проводимости, то получим эту же запись. Точно такую же форму можно получить из сопротивлений для параллельной схемы замещения:

$$\underline{Z} = R \cdot jX_L/(R + jX_L) = R \cdot X_L^2/(R^2 + X_L^2) + jX_L \cdot R^2/(R^2 + X_L^2) = r + jx_L. \qquad (4)$$

Вариант преобразования с использованием понятия о добротности

Метод, обозначенный выражениями (1) и (2), широко используется во всех учебных пособиях. Основной недостаток — наличие в одном выражении величин с неодинаковой размерностью и невозможность или скорее сложность сравнения, например, r и R и g и G для любой пары преобразуемых схем. Понятие о добротности катушки индуктивности или конденсатора обычно вводят радиотехники, которые их связывают с добротностью колебательного контура. Не вдаваясь в подробности [2], представим их отношениями тех или иных параметров рассматриваемого элемента. Например, добротность катушки индуктивности может быть найдена из отношений сопротивлений или проводимостей для каждой из схем по-разному $Q_L = x_L/r = g/b_L = B_L/G = R/X_L$ и для конденсатора: $Q_C = x_C/r = g/b_C = B_C/G = R/X_C$, т. е. одинаково для обоих вариантов схем замещения рассматриваемых устройств. С учетом записанных отношений параметров выражения (3) и (4) можно переписать иначе:

$$G = g \cdot \frac{b_L^2}{g^2 + b_L^2} = \frac{g}{1 + Q_L^2};\ \text{откуда}\ R = r \cdot (1 + Q_L^2). \qquad (5)$$

Или $B_L = b_L \cdot \frac{g^2}{g^2 + b_L^2} = \frac{b_L}{1 + D_L^2}$; откуда

$$X_L = x_L \cdot (1 + D_L^2); \qquad (6)$$

где $D_L = 1/Q_L$ – затухание, величина обратная добротности.

Достоинство выражений (5) и (6) состоит в том, что из них даже без чисел видно, что при достаточно большой добротности (больше десяти) величины реактивных сопротивлений обоих схем практически равны, активные же сопротивления двух различных структур тем больше различны по величине, чем лучше добротность.

Вариант преобразования по закону дуальных структур

Закон дуальных структур [3] для каждой из пар рассматриваемых схем очевиден и из выражений (1) и (2), достаточно перейти от G к R или от B к X и выполнить несложные преобразования, связанные с получением произведений (7):

$$r \cdot R = Z^2 \quad и \quad x \cdot X = Z^2. \tag{7}$$

Эти соотношения в равной мере применимы для каждой из пар и очень удобны для использования в преобразованиях рассматриваемых структур цепей, так как совершенно равноценны для любой из двух связанных эквивалентностью схем. Однако заметим, что выражения (5) и (6) все же более практичны чем (7), так как определить значение $Z = \sqrt{r^2 + x^2}$ более затруднительно, чем найти значение добротности $Q = x / r$.

Частотные преобразования

Кроме того, в практике анализа RC и RL-цепей обычно работают с RLC-параметрами. В этой работе использованы понятия об активных и реактивных сопротивлениях из очень простых соображений. Пусть известно, что реактивное сопротивление катушки $x = 2\pi fL = wL = \frac{ww_0L}{w_0} = \frac{w \cdot x_0}{w_0} = \omega \cdot x_0$ и на базовой частоте имеет конкретное значение $x_0 = 40\ Ом$. Чтобы сделать его частотно зависимым достаточно умножить на безразмерный комплексный оператор $j\omega$. Например, $x(\omega) = 40j\omega$. Здесь $\omega = w / w_0$ — относительная круговая частота; w – круговая частота; w_0 – круговая базовая частота. Другими словами, умножая индуктивное сопротивление на относительную угловую частоту, практически ничего не меняя по существу, обеспечивают переход от частоты $w = 2\pi f$, имеющей размерность рад/с к безразмерной относительной частоте, которая на базовой частоте всегда равна единице. Эта же операция может быть проделана с емкостным сопротивлением или проводимостью. Пусть $x = \frac{1}{jwC} = \frac{1}{\frac{jw \cdot w_0C}{w_0}} = \frac{x_0}{j\omega}$; тогда при $x_0 = 40\ Ом$, имеем $x(\omega) = 40/(j\omega)$.

Авторы рассчитывают, что этот удобный переход от циклической частоты к относительной позволит без труда решать многие проблемы теории электрических цепей в частотной области с помощью вычислительной техники.

Рассмотрим сопротивления четырех цепочек, изображенных на рис. 1, в зависимости от частоты:

$$Z_1(\omega) = r + j\omega \cdot x_0; \quad Y_2(\omega) = \frac{1}{r + j\omega \cdot x_0}; \quad Y_3(\omega) = \frac{1}{r + \frac{x_0}{j\omega}}; \quad Z_4(\omega) = r + \frac{x_0}{j\omega}. \tag{8}$$

Чтобы увидеть частотную зависимость составляющих сопротивлений параллельных схем замещения, необходимо выполнить ряд простых преобразований.

$$Y_2(\omega) = \frac{r - x_0 j\omega}{r^2 + x_0^2\omega^2} = g\frac{r^2}{r^2 + x_0^2\omega^2} - j\omega \cdot b\frac{x_0^2}{r^2 + x_0^2\omega^2} = \frac{G}{\omega^2} \cdot \frac{1+Q^2}{1/\omega^2 + Q^2} + \frac{B\omega}{j} \cdot \frac{1+D^2}{\omega^2 + D^2}. \tag{9}$$

Если учесть, что $\frac{1+D^2}{\omega^2 + D^2} = \frac{1+Q^2}{1+Q^2\omega^2} = \frac{1}{\omega^2} \cdot \frac{1+Q^2}{1/\omega^2 + Q^2} = \frac{M(\omega)}{\omega^2}$ то $Y_2(\omega) = \left(\frac{G}{\omega^2} + \frac{B}{j\omega}\right) \cdot M(\omega)$. Таким образом, параллельная модель линейной цепи в частотном варианте отличается от расчетов на одной частоте двумя коэффициентами

ω^2 и $M(\omega)=\dfrac{1+Q^2}{1/\omega^2+Q^2}$. Если первый коэффициент появляется только у активной составляющей и должен учитываться всегда, то второй учитывают только при добротности $Q \le 10$. При рассмотрении проводимости схемы «в»

$$Y_3(\omega)=G\omega^2\frac{1+Q^2}{\omega^2+Q^2}+j\omega B\frac{1+Q^2}{\omega^2+Q^2}=\left(G\omega^2+j\omega B\right)\cdot N(\omega). \qquad (10)$$

Здесь появляются две частотно зависимые составляющие ω^2 и $N(\omega)=\dfrac{1+Q^2}{\omega^2+Q^2}$. Частотно зависимые коэффициенты $N(\omega)$ и $M(\omega)$ представлены в виде графических зависимостей на рис. 2, из которых видно, что их роль с повышением добротности исчезает. Условно можно принять за граничную единицу $Q \approx 10$.

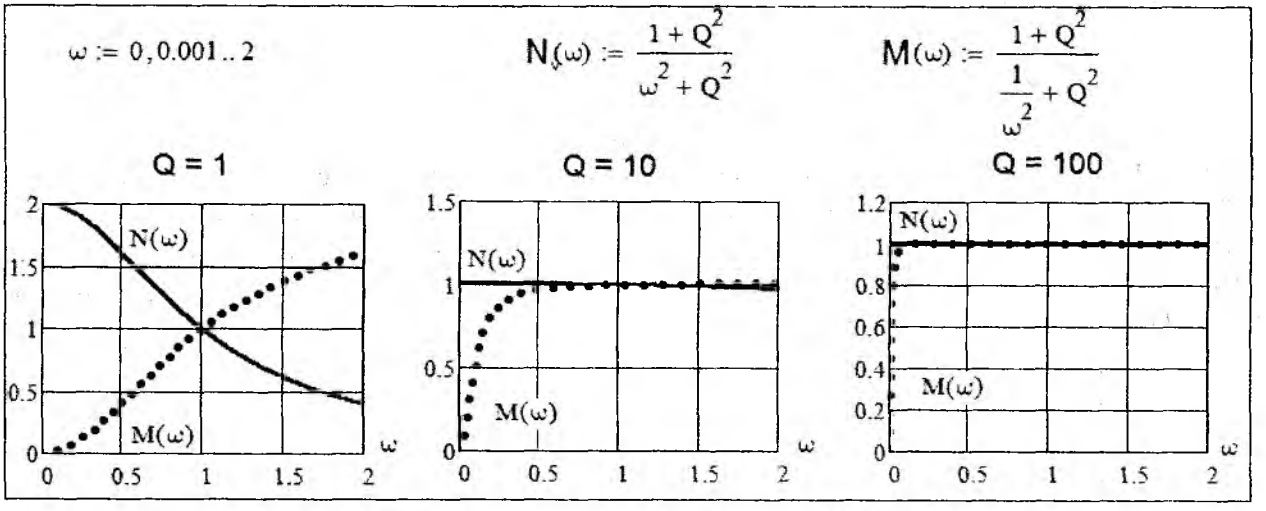


Рис. 2. Частотно зависимые коэффициенты $N(\omega)$ и $M(\omega)$

В практике анализа электрических цепей с конденсаторами и катушками индуктивности возможны случаи, когда активные части проводимостей той и другой цепочки независимы от частоты. В этом случае допустима другая ориентация на эквивалентность. Например, проводимость цепи «в» описана конкретным выражением: $Y_3(\omega)=G+j\omega B$, тогда очевидно что, входное сопротивление будет иметь вид

$$Z_3(\omega)=\frac{1}{G+j\omega B}=R\frac{G^2}{G^2+\omega^2B^2}-j\omega X\frac{B^2}{G^2+\omega^2B^2}=\frac{r}{\omega^2}\frac{1+Q^2}{1/\omega^2+Q^2}-j\omega\cdot x\frac{1+D^2}{\omega^2+D^2}=$$

$$=\left(\frac{r}{\omega^2}+\frac{x}{j\omega}\right)\cdot M(\omega);\quad \textit{Так как}\quad \frac{1+D^2}{\omega^2+D^2}=\frac{1}{\omega^2}\frac{1+Q^2}{1/\omega^2+Q^2}$$

Аналогичный вариант возможен при переходе от параллельной схемы «б» к схеме «а». В отношении базовой частоты оба варианта преобразований идентичны. Что касается боковых частот, то существует некоторая разница в трактовке результата. Однако при добротностях, приближенных к 100 и выше эта разница в полосе пропускания исчезает. Основой для решения служат следующие обстоятельства практики. Катушка индуктивности, как правило, представляется в виде последовательной схемы замещения и в чрезвычайно редких случаях анализа представляется параллельной моделью. Конденсатор же наоборот чаще представляют в виде параллельной модели, которая хорошо соответствует саморазряду. По этой причине предложены два различных алгоритма преобразований рассматриваемых электротехнических устройств.

Катушка индуктивности

За основу принята эквивалентность входных сопротивлений двух схем. Алгоритм состоит из следующей последовательности действий:

- Заданы параметры входного сопротивления, последовательной схемы замещения
- $Z_1(\omega) = r + j\omega \cdot x$.
- Производят расчет вспомогательных параметров и параметров второй схемы на базовой частоте
- $Q = \frac{x}{r}; \quad M(\omega) = \frac{1+Q^2}{1/\omega^2 + Q^2}; \quad G = \frac{1}{r(1+Q^2)}; \quad B = \frac{1}{x(1+D^2)}$.
- Записывают выражение для входной проводимости параллельной цепи:
- $Y_2(\omega) = \left(\frac{G}{\omega^2} + \frac{B}{j\omega}\right) \cdot M(\omega)$.
- Определяют входное сопротивление по выражению
- $Z_2(\omega) = \frac{1}{Y_2(\omega)} = r + j\omega \cdot x$.

На рис. 3 представлен вариант пересчета параметров последовательной схемы замещения катушки индуктивности в эквивалентную ей параллельную схему в соответствии с изложенным алгоритмом при добротности $Q < 10$. На рисунке приведены частотные зависимости модуля и фазы входного сопротивления совершенно аналогичные для обеих схем. Кроме того даны значения активной и реактивной составляющих этого сопротивления для трех значений относительных частот.

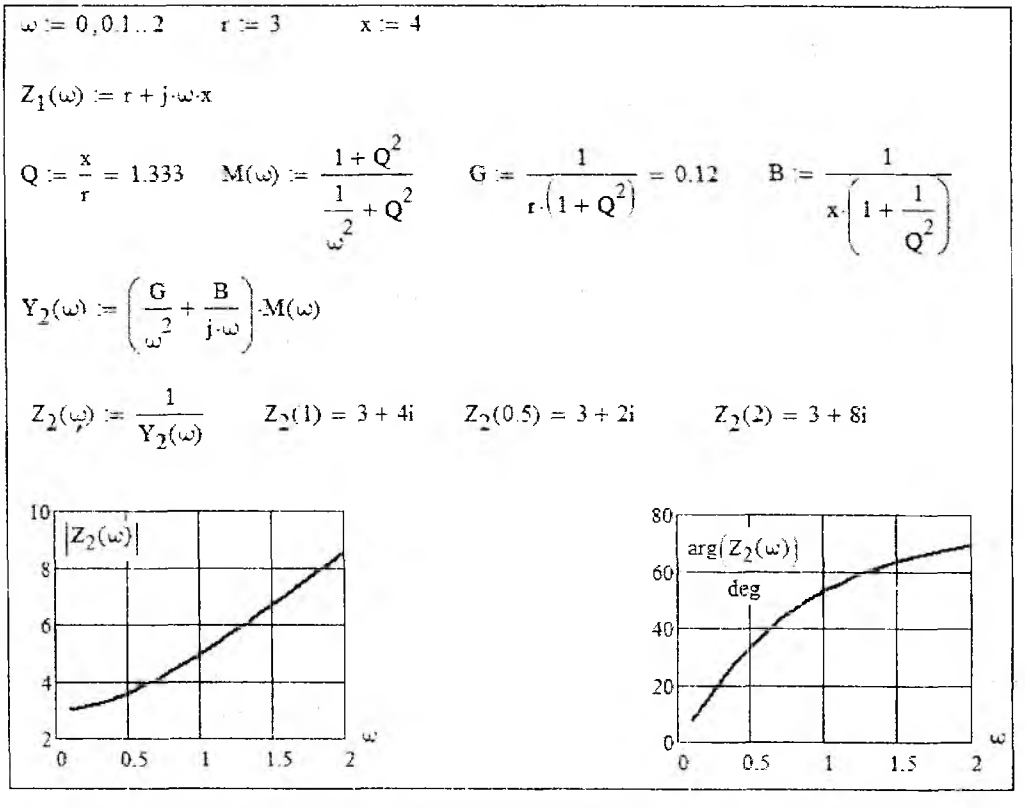


Рис. 3. Пример эквивалентного пересчета последовательной схемы в параллельный аналог

Конденсатор

За основу принята эквивалентность входных проводимостей двух схем. Алгоритм состоит из следующей последовательности действий:

- Заданы параметры входной проводимости параллельной схемы замещения

- $Y_2(\omega) = G + j\omega \cdot B$.
- Производят расчет вспомогательных параметров и параметров второй схемы замещения на базовой частоте
- $$Q = \frac{B}{G}; \quad r = \frac{1}{G(1+Q^2)}; \quad x = \frac{1}{B(1+D^2)}; \quad M(\omega) = \frac{1+Q^2}{1/\omega^2 + Q^2}$$
- Записывают выражение для входного сопротивления последовательной цепи
- $$Z_1(\omega) = \left(\frac{r}{\omega^2} + \frac{x}{j\omega}\right) \cdot M(\omega)$$
- Определяют входную проводимость по известному выражению
- $Y_1(\omega) = \dfrac{1}{Z_1(\omega)} = G + j\omega \cdot B$.

Ниже, на рис. 4, представлен вариант анализа схем замещения конденсатора в соответствии с описанным выше алгоритмом.

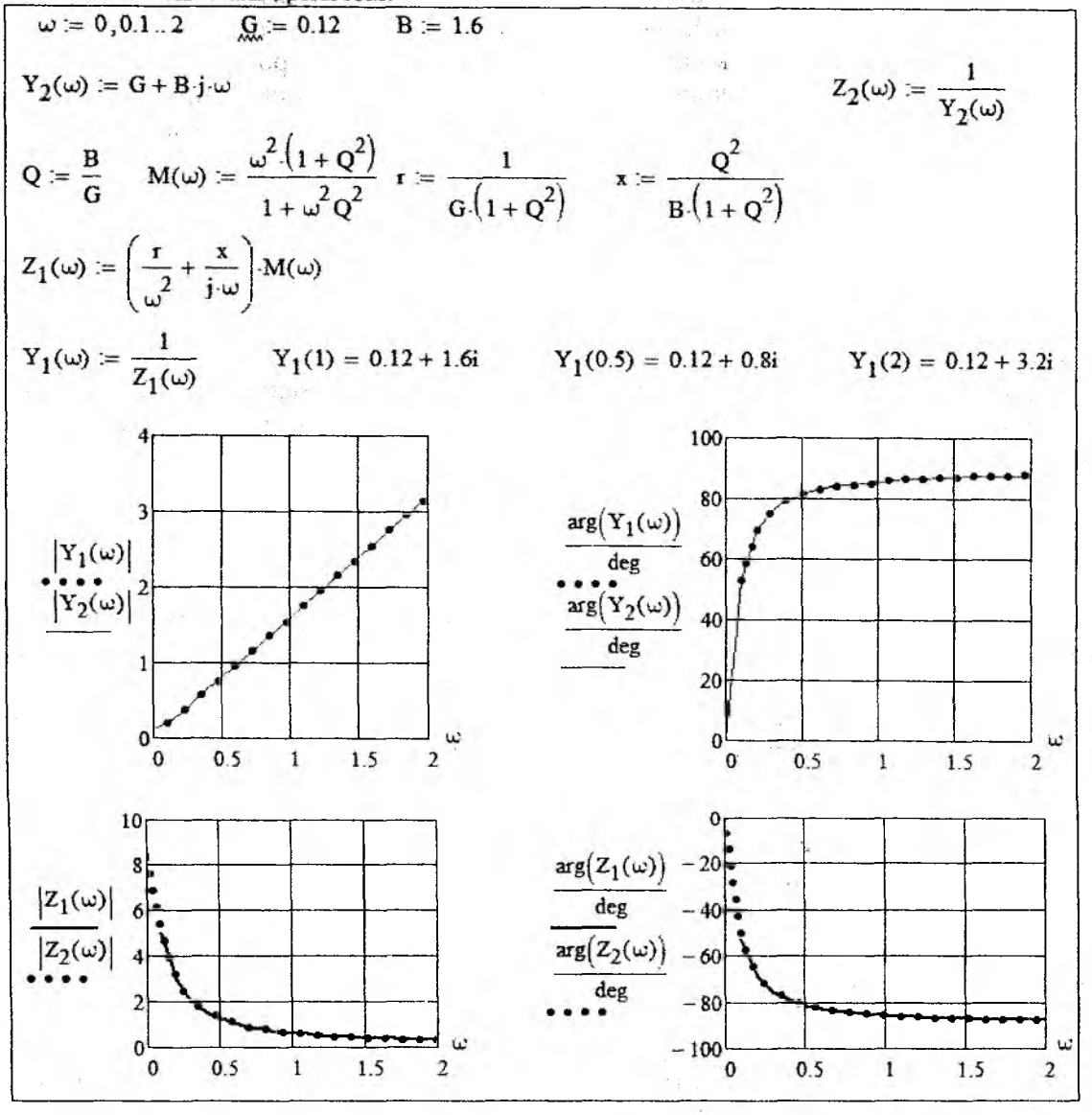


Рис. 4. Пример эквивалентного пересчета параллельной схемы в последовательный аналог

Заметим, что в данном примере частотно-зависимый коэффициент $M(\omega)$ можно было опустить из рассмотрения. Здесь он равен $M(0.5) = 0.983$ *и* $M(2) = 1.004$ при добротности $Q = 13.3$.

Последовательный колебательный контур

Его обычно рассматривают в составе из трех параметров (RLC), пренебрегая параллельным сопротивлением в схеме конденсатора. В связи с изложенной теорией сопротивление конденсатора в последовательной схеме замещения содержит коэффициент, обратно пропорциональный квадрату относительной частоты при любой добротности.

Такая трактовка схемы замещения конденсатора позволяет учесть и это влияние на резонансный эффект. Так передаточная функция последовательного колебательного контура по напряжению (реакция конденсатора на входное гармоническое воздействие) может быть записана в виде

$$H(\omega)=\frac{\dfrac{1}{G+j\omega\cdot B}}{r+j\omega\cdot x+\dfrac{1}{G+j\omega\cdot B}}=\frac{1}{1+rG(1+j\omega\cdot Q_L)(1+j\omega\cdot Q_C)}=\frac{R}{R+r(1+j\omega\cdot Q)^2}.$$

Последняя форма записи выполнена для варианта с равными добротностями катушки индуктивности и конденсатора. В любом случае, очевидно, что роль активного сопротивления катушки и проводимости конденсатора в колебательном процессе одинаковы – и тот, и другой параметр с соответствующим изменением расширяют полосу пропускания.

Обратить внимание: параллельный контур, составленный из этих же двух схем, описывает процесс изменения тока в индуктивности точно такой передаточной функцией по току:

$$T(\omega)=\frac{\dfrac{1}{r+j\omega\cdot x}}{G+j\omega\cdot B+\dfrac{1}{r+j\omega\cdot x}}=\frac{1}{1+rG(1+j\omega\cdot Q_L)(1+j\omega\cdot Q_C)}.$$

Таким образом, установлено, что процесс резонансных колебаний напряжения на конденсаторе последовательного контура идентичен процессу изменения тока в катушке индуктивности параллельного контура. Можно сделать вывод, что реакция изменения напряжения на катушке в последовательном контуре окажется одинаковой с реакцией общего тока конденсатора по отношению к току в контуре.

Выводы

1. Схемы замещения катушек индуктивности и конденсаторов могут быть рассчитаны эквивалентными и в частотной области или по входному сопротивлению или по входной проводимости, которые в рамках своего алгоритма подчиняются закону соотношений между сопротивлениями и проводимостями.

2. Для катушек индуктивности предпочтительней является эквивалентность по входному сопротивлению, а для конденсаторов — эквивалентность по входной проводимости.

3. При добротности катушек индуктивности или конденсаторов со значением около 100 и выше граница между двумя вариантами пересчета эквивалентности стирается.

4. Частотно-зависимые коэффициенты $M(\omega)$ и $N(\omega)$ при добротности $Q\geq 10$ практически полностью утрачивают влияние на параметры рассчитываемых схем.

5. Последовательный и параллельный колебательные контуры, образованные из схем замещения имеют одинаковые передаточные коэффициенты по току в катушке и по напряжению на конденсаторе.

Список литературы: 1. *Добротворский, И. Н.* Теория электрических цепей : учебник для техникумов. – М. : Радио и связь, 1989. – 472 с. 2. *Новиков, Ю. Н.* Электротехника и электроника. Теория цепей и сигналов, методы анализа : учеб. пособие. – СПб. : Питер, 2005. – 384 с. 3. *Эквивалентные* структуры в гармоническом анализе / П. Ф. Лебедев // Радиотехника. – 2004. – Вып. 139. – С. 107 – 112.

Харьковский национальный
университет строительства и архитектуры

Поступила в редколлегию 04.09.2012

УДК 621.396.677.71

Д.Ю. ПЕНКИН, Л.П. ЯЦУК, д-р физ.-мат. наук

АНАЛИЗ ЭНЕРГЕТИЧЕСКИХ ХАРАКТЕРИСТИК ПОПЕРЕЧНОЙ ЩЕЛИ В ШИРОКОЙ СТЕНКЕ ПРЯМОУГОЛЬНОГО ВОЛНОВОДА С ЛОКАЛЬНЫМ ДИЭЛЕКТРИЧЕСКИМ ВКЛЮЧЕНИЕМ

Введение

Задача дифракции волны основного типа H_{10} на узкой продольной щели, прорезанной в широкой стенке бесконечного прямоугольного волновода над диэлектрической вставкой конечной длины, была решена в работе [1], на поперечной щели – в работе [2]. В обоих случаях в решении задач для определения электрического поля в щели использовался метод Галеркина. При этом электромагнитное поле внутри вставки с диэлектрической проницаемостью ε_1 под щелью находилось методом собственных волн с учетом отражений от торцевых сечений, совпадающих с границами раздела «диэлектрик-воздух» [3]. Тестовые расчеты для поперечной щели в [2, 4] позволили обнаружить возможность получения коэффициента её излучения, близкого к единице, что, согласно теории волноводно-щелевых излучателей [3], недостижимо в случае однородного заполнения волновода диэлектриком. Этот результат стимулировал более детальный численный анализ энергетических характеристик волноводно-щелевого излучателя рассматриваемого типа при варьировании геометрических и электрических параметров диэлектрического включения.

Цель работы – исследование возможностей такого способа управления элементами матрицы рассеяния поперечной щели, излучающей в свободное полупространство или в другой электродинамический объем.

Описание задачи для численного моделирования

Постановка электродинамической задачи, а также её решение для поперечной щели, излучающей в свободное полупространство, представлены в работе [2]. Здесь эта задача обобщена на случай, когда внешней областью для щелевого элемента может быть не только свободное полупространство, но и другой прямоугольный волновод. Причём внешний волновод так же, как и питающий, может содержать локальное диэлектрическое включение с ε_2 (предполагается, что магнитные проницаемости вставок $\mu_1 = \mu_2 = 1$). Заметим, что для волноводной секции (рис.1) задача определения коэффициентов рассеяния нагруженной щелевой неоднородности на первом этапе сводится к определению амплитуд V_{1q} и V_{2q} пространственных гармоник функций распределения электрического поля на поверхностях щели S_1 и S_2, прилегающих к связываемым электродинамическим объемам. Указанные амплитуды определяются из системы (4) линейных алгебраических уравнений (СЛАУ) работы [2]. В случае щелевой связи двух прямоугольных волноводов отличие в матричных элементах СЛАУ [2] сводится к тому, что в качестве внешних проводимостей $Y^e_{22,pq}$ следует использовать проводимости $Y^i_{11,pq}$ [2]. Если волноводы не идентичны, как на рис. 1, $Y^e_{22,pq} = Y^i_{22,pq}$, где $Y^i_{22,pq}$ – внутренняя проводимость щели в верхнем волноводе.

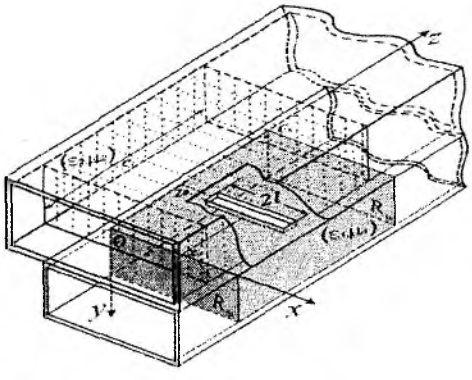


Рис.1. Геометрия волноводно-щелевой секции

При этом может рассматриваться связь волноводов разных поперечных сечений с различными параметрами

диэлектрических включений. Энергетические параметры щелевой неоднородности в этом случае определяются четверкой коэффициентов матрицы рассеяния. Это коэффициенты отражения S_{11} и прохождения S_{12} в основном волноводе (которые находятся, как и в [2]), а также коэффициенты S_{13} и S_{14} во внешнем волноводе, которые представляют собой амплитуды волн, распространяющихся от щели в направлениях $(-z)$ и $(+z)$ соответственно. Используя обозначения работы [2], представим выражения для коэффициентов S_{13} и S_{14}:

$$S_{13}=\sum_{q=1}^{Q}\frac{4V_{2q}U_1(q)}{i\omega\mu_2 bd\pi}\sin\left(\frac{\gamma_{10}^{\varepsilon}d}{2}\right)e^{-i\gamma_{10}^{\varepsilon}z_0}\frac{\left(\gamma_{10}^{\varepsilon}-\gamma_{10}\right)-\left(\gamma_{10}^{\varepsilon}-\gamma_{10}\right)e^{-2i\gamma_{10}^{\varepsilon}(c-z_0)}}{\left(\gamma_{10}^{\varepsilon}+\gamma_{10}\right)^2-\left(\gamma_{10}^{\varepsilon}-\gamma_{10}\right)e^{-2i\gamma_{10}^{\varepsilon}c}} \quad (1)$$

$$S_{14}=\frac{4e^{-i\left(\gamma_{10}^{\varepsilon}-\gamma_{10}\right)c}}{\left(\gamma_{10}^{\varepsilon}+\gamma_{10}\right)^2-\left(\gamma_{10}^{\varepsilon}-\gamma_{10}\right)^2e^{-2i\gamma_{10}^{\varepsilon}c}}\times$$
$$\times\sum_{q=1}^{Q}\frac{4V_{2q}U_1(q)}{i\omega\mu_2 bd\pi}\sin\left(\frac{\gamma_{10}^{\varepsilon}d}{2}\right)\left[\left(\gamma_{10}^{\varepsilon}-\gamma_{10}\right)e^{-i\gamma_{10}^{\varepsilon}z_0}-\left(\gamma_{10}^{\varepsilon}-\gamma_{10}\right)e^{-i\gamma_{10}^{\varepsilon}z_0}\right], \quad (2)$$

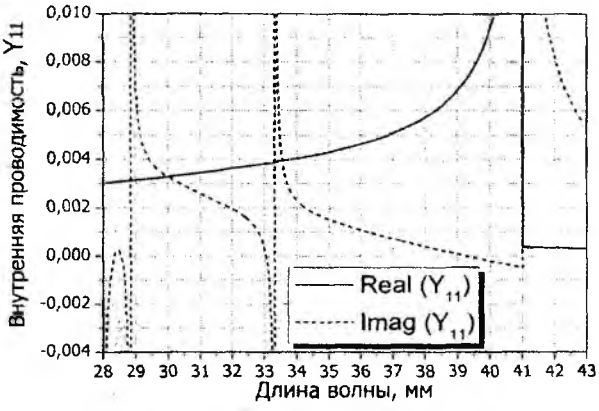
в которых нужно применять электрические и геометрические параметры для внешнего волновода (рис. 1). Заметим, что выражения (1) и (2) получены в предположении одномодового режима работы связанных волноводов. Уравнение баланса мощности здесь имеет следующий вид: $|S_{11}|^2+|S_{12}|^2+|S_{13}|^2+|S_{14}|^2=1$. Для удобства дальнейшего анализа коэффициент передачи мощности во внешний волновод $|S_{\Sigma}|^2=|S_{13}|^2+|S_{14}|^2$, будем обозначать тем же символом, что и коэффициент излучения щели.

Характер диапазонных зависимостей проводимости щели и её энергетических параметров

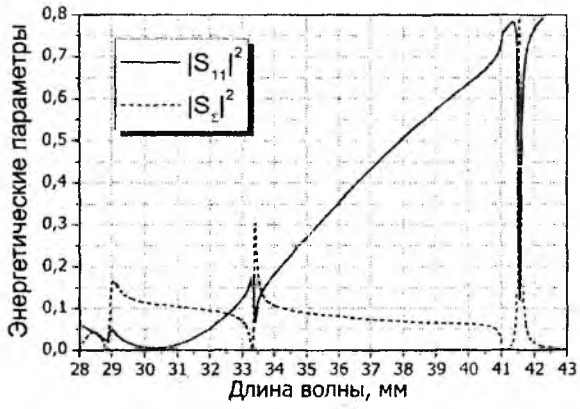
На основании построенных математических моделей для волноводно-щелевого излучателя (рис.1 из [2]) и волноводно-щелевой секции (рис.1) были созданы вычислительные программы на языке Visual Fortran. При расчетах внутренних проводимостей щели Y_{pq}^{i} максимальные значения индексов суммирования m,n в формуле (10) [2] фиксировались равными 20. Число базисных функций Q в аппроксимации (3) [2] выбиралось равным от 1 до 5, толщина стенки волновода полагалась $h=1\,мм$.

Известно [1], что в полосе частот одномодового режима работы полого волновода внутри диэлектрической вставки могут возбуждаться распространяющиеся высшие моды, так называемые «запертые моды». По мере увеличения значения ε_1 их критические частоты приближаются к критической частоте основной волны полого волновода. Это следует учитывать при выборе значений диэлектрической проницаемости вставки. На критических частотах значение $\gamma_{mn}^{\varepsilon}=\sqrt{k^2\varepsilon_1-(m\pi/a)^2-(n\pi/b)^2}$ для этих волн обращается в ноль. При этом одно из слагаемых в выражении для внутренних проводимостей (10) [2] стремится к бесконечности, и, как следствие, поперечная щель не излучает. Кроме того, коэффициенты отражения $R_{-\mu}^{э(м)}$ и $R_{\mu}^{э(м)}$ для «запертых мод» по абсолютной величине равны единице. Поэтому на одной из частот, на которой $y_{mn}^{\varepsilon}c=v\pi$, может оказаться, что знаменатель $(1-R_{\mu}^{э(м)}R_{\mu}^{э(м)})$ в выражении (10) [2] обращается в ноль, а соответствующее слагаемое – в бесконечность.

В этом случае щель также оказывается неизлучающей. В результате зависимость коэффициента излучения (равно как и других коэффициентов матрицы рассеяния волноводно-щелевого излучателя) от частоты носит очень сложный изрезанный характер подобно тому, как это наблюдалось и в работе [1]. Для иллюстрации на рис.2 приведены частотные характеристики волноводно-щелевого излучателя с параметрами: a=23*мм*, b=10*мм*, $x_0 = a/2$, c=7*мм*, $z_0 = 0.6$*мм*, $\varepsilon = 5$, $2l = 9.15$*мм* и d=1*мм*. Ближайшим высшим типом волны, возбуждаемым щелью, расположенной симметрично относительно оси широкой стенки волновода, является волна типа H_{11}. Для волновода стандартного поперечного сечения 23x10*мм* при $\varepsilon = 5$ одномодовый режим в области вставки поддерживается при $\lambda > 41.01$*мм*.



а - Реальная и мнимые части внутренней проводимости Y_{11}



б - Диапазонная зависимость энергетических параметров щели

Рис.2. Диапазонные характеристики волноводно-щелевого излучателя при a=23мм, b=10мм, $x_0 = a/2$, c=7мм, $z_0 = 0.6$ мм, $\varepsilon = 5$, $2l = 9.15$ мм и d=1мм

Как видно из рис.2, *б*, в коротковолновой части диапазона, где существует многомодовый режим работы диэлектрической вставки, наблюдаются участки квазипостоянного уровня коэффициента излучения щели при малой величине коэффициента отражения (например, при $\lambda \in [29.5 \div 31.5]$). В отличие от одномодового режима уровни коэффициента излучения оказываются достаточно низкими, что раскрывает возможность согласования на входе резонансной волноводно-щелевой решетки из таких элементов. Однако управление характеристиками одиночной щели проще обеспечить в части диапазона одномодового режима для вставки, где частотные зависимости являются более монотонными.

Энергетические параметры волноводно-щелевого излучателя в одномодовом режиме диэлектрической вставки

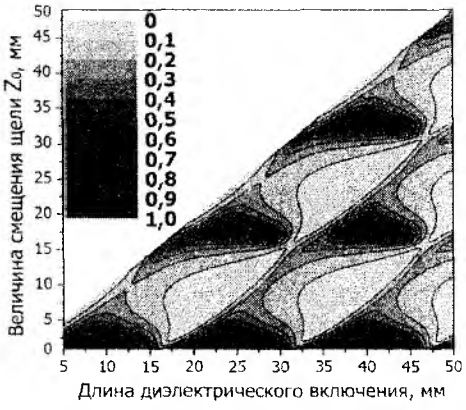
Известно, что поперечная щель, в отличие от продольной, имеет ограниченные возможности изменения величины $|S_\Sigma|^2$ за счет своего смещения относительно оси волновода. При наличии вставки появляется возможность управления коэффициентами отражения $|S_{11}|^2$ и излучения $|S_\Sigma|^2$ волноводно-щелевого излучателя путем изменения длины вставки при фиксированном значении ε и положения щели над вставкой. Для диэлектрической вставки с $\varepsilon = 3$ было установлено, что подбором геометрических параметров рассматриваемой неоднородности можно изменять величину коэффициента излучения резонансной щели в пределах от нуля до значения, близкого к единице.

Показательно, что значение коэффициента $|S_\Sigma|^2$, близкое к единице, удается получить вблизи критической длины волны в полом волноводе. В подтверждение этого на рис. 3 в ра-

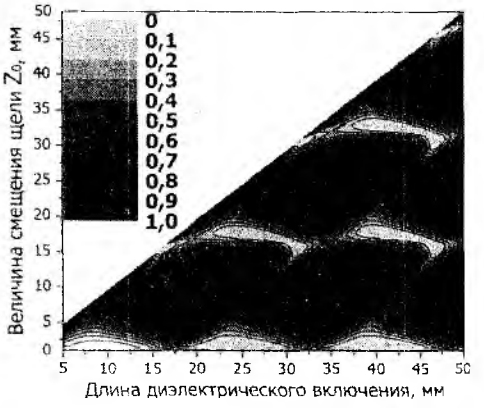
боте [2] приведены диапазонные характеристики волноводно-щелевого излучателя с параметрами: a=23$мм$, b=10$мм$, $x_0 = a/2$, c=10.5$мм$, $z_0 = 0.75 мм$, $\varepsilon = 3.0$, $2l = 16 мм$ и d=1$мм$.

На частоте $f = 6.74\,ГГц$ коэффициент $|S_\Sigma|^2$ достигает значения 0.94 (здесь длина волны $\lambda = 44.5 мм$, длина волны в волноводе с однородным диэлектрическим заполнением $\lambda_в^\varepsilon \approx 31 мм$). Заметим, что в этом случае резонансная длина щели оказалось равной $0.55\lambda_в^\varepsilon$.

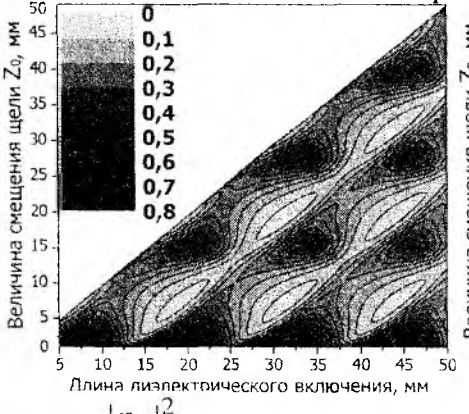
Диэлектрическое включение способно обеспечить увеличение уровня коэффициента излучения $|S_\Sigma|^2$ за счет взаимного погашения волн, возбужденных щелью и отраженных от диэлектрической вставки (а также прошедших за вставку), что приводит к уменьшению коэффициентов отражения $|S_{11}|^2$ и прохождения $|S_{12}|^2$. Возможно также и полное прохождение падающей волны за неоднородность «щель-диэлектрик». В этом случае волна, прошедшая за вставку, не гасится волной, возбужденной щелью, а напротив – усиливается ею. Другими словами, существует возможность реализации окна прозрачности для этой системы, что и наблюдается на рис.3 в [2] в окрестности частоты $f = 9.0\ ГГц$.



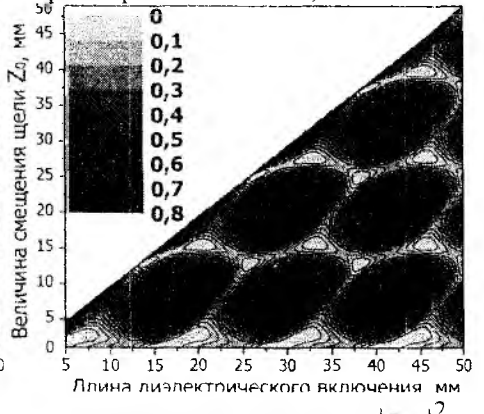
а - Коэффициент излучения $|S_\Sigma|^2$ при параметрах: $2l$ = 16 *мм*, λ=44.5 *мм*.



б - Коэффициент отражения $|S_{11}|^2$ при параметрах: $2l$ = 16 *мм*, λ=44.5 *мм*.



в - Коэффициент излучения $|S_\Sigma|^2$ при параметрах: $2l$ = 13 *мм*, λ=38 *мм*.



г - Коэффициент излучения $|S_{11}|^2$ при параметрах: $2l$ = 13 *мм*, λ=38 *мм*.

Рис.3. Влияние длины вставки и смещения щели Z_0 на энергетические параметры волноводно–щелевого излучателя при a=23 *мм*, b=10 *мм*, $x_0 = a/2$, $\varepsilon = 3.0$, $2l = 16$ *мм* и d=1 *мм*

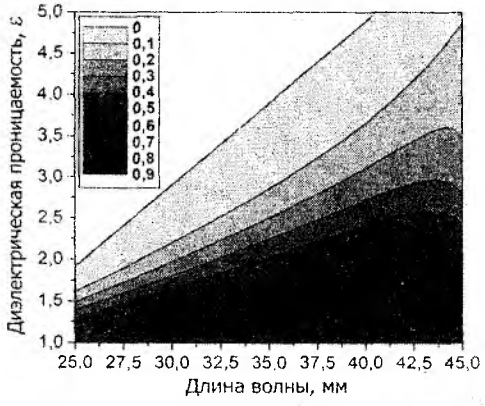
Энергетические параметры волноводно-щелевого излучателя согласно выражениям, приведенным в [2], являются периодическими функциями аргументов c и z_0. Эта периодичность была численно исследована на фиксированной длине волны $\lambda = 44.5 мм$ при

геометрических размерах щели $2l = 16$ *мм*, d=1*мм* и a=23*мм*, , b=10*мм*, $x_0 = a/2$, $\varepsilon = 3.0$. Двумерные зависимости при изменении $c \in [5;50]$ *мм* и $z_0 \in [0.6; c-0.6]$ *мм* для коэффициентов излучения $|S_\Sigma|^2$ и отражения $|S_{11}|^2$ приведены на рис.3, *а, б* соответственно (в расчетах интервал для параметра z_0 был выбран исходя из требования неравенства $d/2 < z_0 < c - d/2$). Как видно из рисунка, зоны максимального излучения и минимального отражения наблюдаются для вставок с продольным размером c, кратным нечетному числу $\lambda_в^\varepsilon/4$ для основной моды. Причем для этих зон расстояние z_0 между центром щели и левой границей вставки должно быть кратно $\lambda_в^\varepsilon/2$. Эти закономерности подтверждаются аналогичными численными расчетами характеристик волноводно-щелевого излучателя и на других частотах (рис.3, *в, г*).

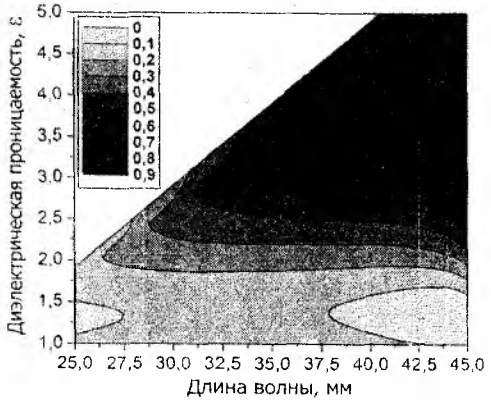
Следует отметить, что выявленные закономерности максимального излучения поперечной щели справедливы только в диапазоне одномодового режима возбуждения диэлектрической вставки. При увеличении значения ε и длины вставки ширина этого диапазона уменьшается. Представляет интерес, как изменяются по частоте энергетические характеристики волноводно-щелевого излучателя с изменением ε при условии, что во всех частотных точках $c = \lambda_в^\varepsilon/4$ и длина щели $2l = [0.471 - 0.046(\varepsilon - 1)]\lambda$. Заметим, что используемое здесь выражение для аппроксимации длины щели было получено на основании предварительной оценки резонансной длины поперечной щели шириной d=1*мм*, прорезанной в бесконечном волноводе сечением 23x10*мм* полностью заполненном диэлектриком, для интервала $31 < \lambda < 33$ *мм*.

Такие характеристики в виде двумерных зависимостей приведены на рис.4, *а,б*. Здесь a=23*мм*, b=10*мм*, $z_0 = 0.6$*мм*, $x_0 = a/2$, d=1*мм*, а область частот при исследовании ограничена критической частотой волны типа H_{11} в диэлектрическом включении. Заметим, что $z_0 = 0.6$*мм* соответствует крайнему левому положению щели относительно диэлектрического включения, при котором наблюдается максимальное значение коэффициента излучения $|S_\Sigma|^2$. Как видно из результатов численного моделирования, условия максимального излучения для щели $|S_\Sigma|^2 \approx 0.9$ реализуются в длинноволновой части рабочего (одномодового) диапазона полого прямоугольного волновода стандартного поперечного сечениям при достаточно малых $\varepsilon \in [1.25;1.65]$. Следует отметить, что при этом коэффициент излучения щели $|S_\Sigma|^2 \geq 0.7$ достижим практически во всех точках рабочего диапазона волновода. Однако, как было ранее показано на рис.3, *а* для случая $\varepsilon = 3.0$, подобные условия излучения наблюдаются для щелей с более короткими длинами. При $2l = 0.55\lambda_в^\varepsilon$ и сохранении остальных геометрических параметров, характеристики щелевого излучателя представлены на рис.4, *в, г*. С увеличением значения диэлектрической проницаемости диапазон частот одномодового режима волновода сужается и смещается к критической длине волны полого волновода. При этом для $\varepsilon \geq 6.29$ одномодовый режим во вставке становится невозможным, поскольку критическая длина волны H_{11} превышает 46 *мм* – критическую длину волны в полом волноводе. В свою очередь, для каждого значения диэлектрической проницаемости из диапазона $1 \leq \varepsilon < 6.29$ существует возможность подобрать длину щели такую, что для нее будет наблюдаться условие максимального излучения $|S_\Sigma|^2 \approx 0.9$. Причем это осуществимо только в областях близких к критической длине волны полого волновода $\lambda_{кр} = 46$*мм*. Например, в случае $\varepsilon = 1.3$ (рис.4, *а*) максимальный коэффициент излучения $|S_\Sigma|^2 \approx 0.9$ достигается на

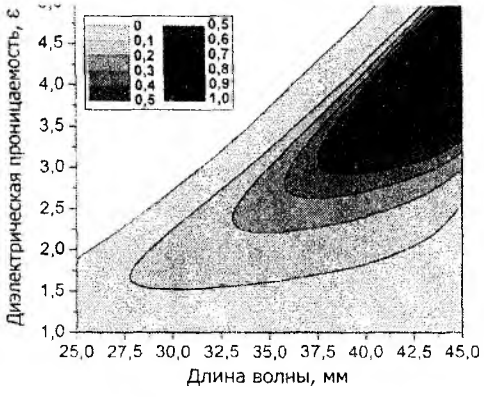
длине волны $\lambda = 0.95\lambda_{кр}$. При укорочении длины волны коэффициент излучения такой щели постепенно уменьшается до уровня $|S_\Sigma|^2 \approx 0.5$ при $\lambda = 26\,мм$.



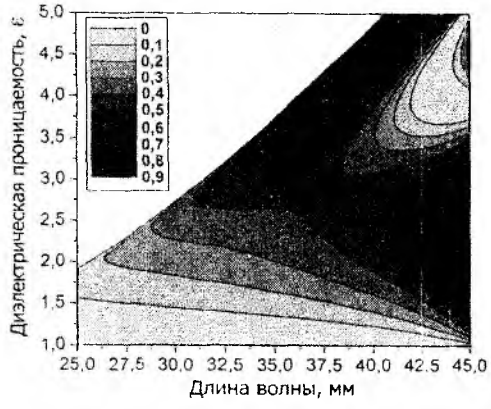
а - Коэффициент излучения $|S_\Sigma|^2$.



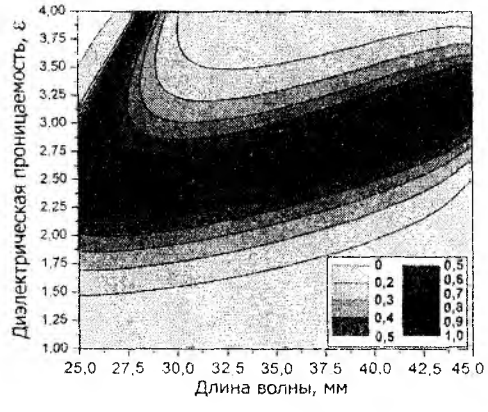
б - Коэффициент отражения $|S_{11}|^2$.



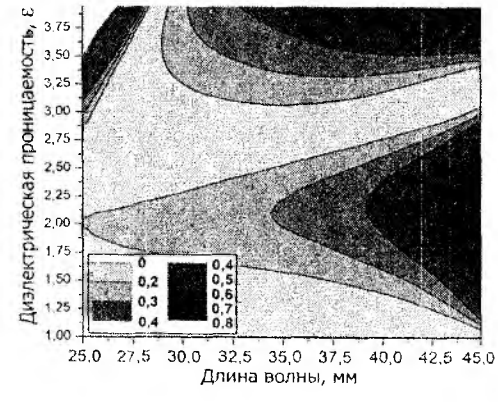
в - Коэффициент излучения $|S_\Sigma|^2$ при параметрах: $2l = 0.55\lambda_в^\varepsilon$, b=10 *мм*.



г - Коэффициент отражения $|S_{11}|^2$ при параметрах: $2l = 0.55\lambda_в^\varepsilon$, b=10 *мм*.



д - Коэффициент излучения $|S_\Sigma|^2$ при параметрах: $2l = \lambda_в^\varepsilon/2$, b=7 *мм*.



е - Коэффициент отражения $|S_{11}|^2$ при параметрах: $2l = \lambda_в^\varepsilon/2$, , b=7 *мм*.

Рис. 4. Энергетические параметры волноводно-щелевого излучателя при условии одномодового режима в области вставки при a=23 *мм*, $c = \lambda_в^\varepsilon/4$, $z_0 = 0.6\,мм$, $x_0 = a/2$, d=1 *мм*

Влияние высоты волновода на диапазонные свойства структуры

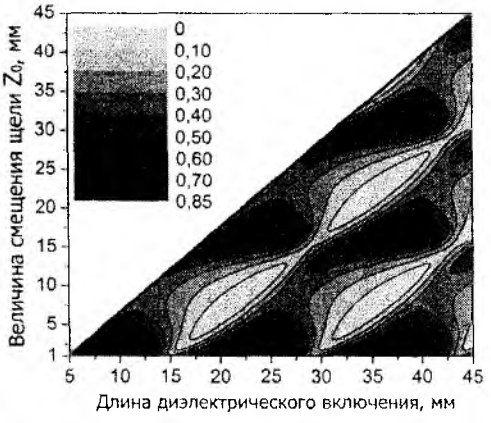
С целью обеспечения условий максимального излучения щели на более коротких длинах волн рабочего диапазона волновода были исследованы заниженные (низкопрофильные) волноводы. Действительно, уменьшение значения высоты волновода b приводит к уменьшению величины критической длины волны типа H_{11} и, следовательно, к расширению одномодового диапазона диэлектрического включения в сторону коротких длин волн. Это позволяет полагать, что для низкопрофильных волноводов условия высокого излучения щели в случаях $2 \le \varepsilon < 6.29$ могут проявляться и для более коротких длин волн, чем в случае волновода стандартного поперечного сечения. Результаты расчетов подтвердили такое предположение.

На рис.4, *д, е* представлены двумерные зависимости энергетических характеристик щели длиной $2l = \lambda_в^\varepsilon / 2$ для высоты волновода b=7*мм* (при a=23*мм*, $c = \lambda_в^\varepsilon / 4$, $x_0 = a/2$ $z_0 = 0.6$*мм*, d=1*мм*), подобные рассмотренным выше диапазонным зависимостям в волноводе стандартного поперечного сечения. Как видно из рисунков для такого заниженного волновода при диэлектрической проницаемости включения $\varepsilon \in [2.3;3.0]$ можно обеспечить значение коэффициента излучения щели $|S_\Sigma|^2 \ge 0.8$ для любой длины волны в интервале рабочего диапазона полого волновода. Результат может быть получен и при дальнейшем занижении волновода с условием использования диэлектрического включения с более высоким значением ε .

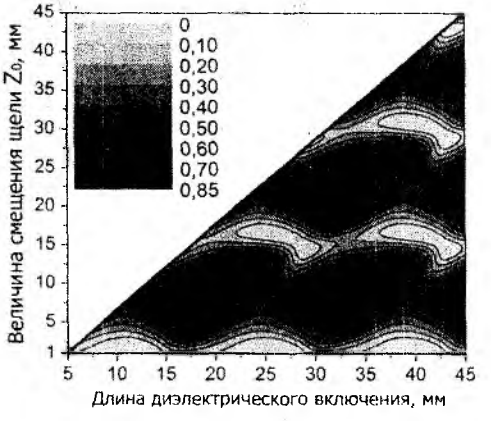
Энергетические параметры волноводно-щелевой секции

Проведенный анализ численных результатов подтвердил возможность управления энергетическими характеристиками волноводно-щелевого излучателя посредством локального диэлектрического включения. Для практических приложений важно доказать с помощью численного моделирования сохранение таких возможностей управления энергетическими параметрами щели и для случаев её излучения в объёмы, отличные от полупространства. В частности, в случае волноводной секции, изображенной на рис.1, получены следующие результаты.

В качестве примера на рис.5, *а, б* представлены двумерные зависимости влияния длины вставки c и смещения щели z_0 на энергетические характеристики волноводно-щелевой секции с параметрами: a=23 *мм*, b=10 *мм*, $\lambda = 42.5$ *мм*, $2l = 13$ *мм*, $\varepsilon = 3.0$ и d=1*мм*.



а - Коэффициент излучения $|S_\Sigma|^2$.

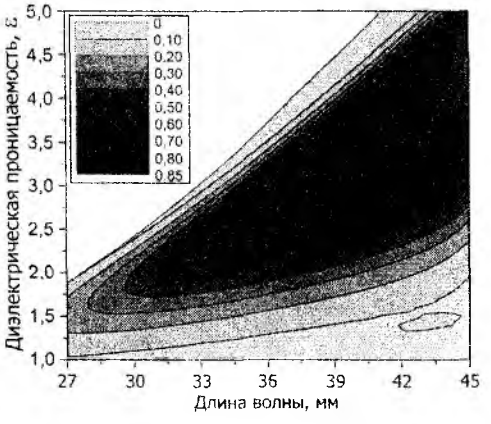


б - Коэффициент отражения $|S_{11}|^2$.

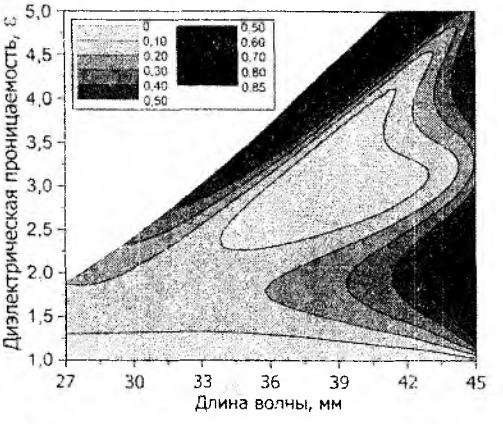
Рис.5. Влияние длины вставки и смещения щели z_0 на энергетические параметры волноводно-щелевого излучателя при a=23 мм, b=10 *мм*. $\lambda = 42.5$ *мм*, $2l = 13$ *мм*, $x_0 = a/2$, $\varepsilon = 3.0$ и d=1 *мм*

Здесь для удобства сопоставления результатов расчетов длина щели выбрана такая же, как и для случая волноводно-щелевого излучателя (рис.3, *в, г*), а длина волны была скорректирована с целью максимизации коэффициента передачи мощности во внешний волновод.

В качестве внешнего здесь рассматривается полый прямоугольный волновод такого же поперечного сечения, как и возбуждаемый волновод с диэлектрическим включением. Поскольку в рассматриваемом случае $|S_{13}|^2 = |S_{14}|^2$, то для удобства анализа на рисунке приведены значения коэффициента $|S_\Sigma|^2 = |S_{13}|^2 + |S_{14}|^2$. Как и в случае излучения щели в полупространство, зоны её максимального излучения и минимального отражения наблюдаются для вставок с продольным размером c кратным нечетному числу $c = \lambda_в^\varepsilon/4$ и расстоянием z_0 между центром щели и левой границей вставки кратным $c = \lambda_в^\varepsilon/2$. Отметим, что условия её максимального излучения здесь также реализуются только в длинноволновой части одномодового рабочего диапазона полых прямоугольных волноводов стандартного поперечного сечения. Расчеты показали, что в случае двух связанных волноводов максимально достижимая передаваемая мощность щели оказывается несколько меньше чем максимальная мощность, излучаемая щелью в полупространство. Например, на длине волны $\lambda = 42.5\,мм$ для щели длиной 13 *мм* во внешний волновод передается меньшая мощность $|S_\Sigma|^2 \approx 0.85$, чем мощность излучения в полупространство $|S_\Sigma|^2 \approx 0.95$ щели длиной 16 *мм* при $\lambda = 44.5\,мм$. Подобная картина сохраняется и при введении диэлектрической вставки во внешний волновод. Однако при этом наблюдается возможность смещения этого уровня в область более коротких длин волн по сравнению со случаем полого внешнего волновода. В качестве примера на рис.6, *а,б* представлены двумерные зависимости энергетических характеристик волноводно-щелевой секции при условии одномодового режима в области идентичных вставок в обоих волноводах с одинаковым поперечным сечением при a=23 *мм*, $c = \lambda_в^\varepsilon/4$ $z_0 = 0.6\,мм$, $x_0 = a/2$, $2l = 0.45\lambda_в^c$, d=1*мм*. Как видно из рисунка, уровень $|S_\Sigma|^2 \approx 0.85$ может быть обеспечен на длине волны $\lambda \approx 39\,мм$ при $\varepsilon = 3$.



а - Коэффициент излучения $|S_\Sigma|^2$ при ширине волновода b = 10 *мм*



б - Коэффициент отражения $|S_{11}|^2$ при ширине волновода b = 10 *мм*

Рис.6. Влияние длины вставки и смещения щели z_0 на энергетические параметры волноводно-щелевого излучателя при a=23мм, $c = \lambda_в^\varepsilon/4$ $z_0 = 0.6\,мм$, $2l = 0.45\lambda_в^\varepsilon$, d=1*мм*

Как показывают расчеты, дальнейшее уменьшение резонансной длины волны волноводно-щелевого устройства можно обеспечить, используя в его конструкции низкопрофильные волноводы.

Выводы

На основании исследований диапазонных характеристик волноводно-щелевого излучателя при различных вариантах его геометрической конфигурации и значениях диэлектрической проницаемости локальной вставки доказана возможность управления в широких пределах величинами его энергетических параметров (коэффициентов излучения и отражения). В частности, установлено, что зоны максимального излучения и минимального отражения наблюдаются для вставок с продольным размером c кратным нечетному числу $\lambda_в^\varepsilon/4$. Причем для этих зон расстояние z_0 между центром щели и левой границей вставки должно быть близко к $d/2$ или кратно $\lambda_в^\varepsilon/2$ (если второе условие может быть физически выполнено). Также выявлено, что условия максимального излучения щели могут быть реализованы только в длинноволновой части одномодового диапазона полого прямоугольного волновода стандартного поперечного сечения. С целью обеспечения условий максимального излучения щели на более коротких длинах волн могут быть использованы заниженные (низкопрофильные) волноводы. Для случая волноводно-щелевой секции (при излучении щели в прямоугольный волновод с диэлектрической вставкой) подтверждены возможности управления энергетическими характеристиками волноводного устройства с помощью изменения параметров диэлектрических включений.

Список литературы: 1. *Яцук, Л.П., Бердник, С. Л., Катрич, В. А.* Энергетические характеристики щели в волноводе с диэлектрической вставкой конечной длины // Радиоэлектроника. – 2007. – №11 – С.44-55. 2. *Яцук, Л.П., Пенкин, Д.Ю.* Дифракция волны типа H_{10} на поперечной щели в широкой стенке прямоугольного волновода с локальным диэлектрическим включением // Радиофизика и радиоастрономия. – 2010. – Т.15, № 4 – С. 434-441. 3. *Фельд, Я.Н, Бененсон, Л.С.* Антенно-фидерные устройства. – М. : Изд-во ВВИА им Н.Е. Жуковского,1959. – Ч.2. – 551с. 4. *Penkin, D.Yu., Yatsuk, L.P.* The Output Performance of Cross Slot in Rectangular Waveguide with Dielectric Insetion of Finite Length // Proc. International Kharkov Symposium on Physics and Engineering of Microwaves, Millimeter and Submillimeter waves, 2010. – A14 – http://www.ire.kharkov.ua/.

Харьковский национальный университет
имени В.Н. Каразина

Поступила в редколлегию 15.09.2012

УДК 621.372.832

С. Л. БЕРДНИК, канд. физ.-мат. наук

ИЗЛУЧЕНИЕ ЭЛЕКТРОМАГНИТНЫХ ВОЛН ЭЛЕКТРИЧЕСКИ ДЛИННОЙ ЩЕЛЬЮ С ДИЭЛЕКТРИЧЕСКИМ ЗАПОЛНЕНИЕМ В УЗКОЙ СТЕНКЕ МНОГОМОДОВОГО ПРЯМОУГОЛЬНОГО ВОЛНОВОДА

Введение

При практической реализации высоких технических характеристик радиотехнических комплексов, решении задач электромагнитной совместимости целесообразным является применение многофункциональных многочастотных излучающих структур. В случае волноводно-щелевых излучателей бегущей волны представляет интерес реализация многочастотной антенны на основе электрически длинной щели в узкой стенке многомодового прямоугольного волновода, возбуждаемого несколькими волноводными модами типа H_{m0} на разных частотах. Но, в отличие от щелевого излучателя в широкой стенке волновода, электродинамические характеристики которого можно оптимизировать путем смещения щели к середине волновода с одновременным увеличением ее ширины [1], для щели в узкой стенке получить такой результат невозможно [2, 3]. Улучшить характеристики направленности таких щелевых излучателей (уменьшить уровень боковых лепестков диаграммы направленности (ДН), увеличить коэффициенты направленного действия (КНД) и усиления (КУ)) можно за счет увеличения толщины стенки волновода [4], реализации спадающего к краям щели амплитудного распределения поля при определенных соотношениях между длиной щели и длинами волн в свободном пространстве и в волноводе [1, 2], а также при использовании диэлектриков в полости щели [2].

Постановка задачи и решение

Рассмотрим продольную щель длиной l и шириной d ($L >> d$, $d < \lambda$, λ – длина волны в свободном пространстве), которая прорезана в узкой стенке конечной толщины h прямоугольного волновода с размерами поперечного сечения $a \times b$. Полость щели заполнена диэлектриком с относительной диэлектрической проницаемостью $\dot{\varepsilon}^{v} = \varepsilon^{v}(1 - i \operatorname{tg} \delta^{v})$ (рис. 1). Волноводно-щелевой излучатель возбуждается волной типа H_{m0} (m=1, 2, 3…).

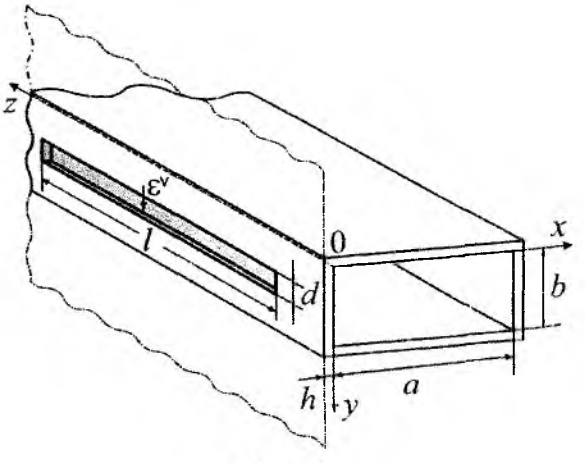


Рис. 1

Учет толщины стенки волновода сводится к решению задачи о связи трех электродинамических объемов: полости волновода (индекс "*in*"), полости щели (индекс "*v*") и внешнего объема (индекс "*ext*"), который в нашем случае представляет собой полупространство над бесконечной идеально проводящей плоскостью. Из условий непрерывности тангенциальных

составляющих магнитного поля на поверхностях щели ς_1 (волновод – щель) и ς_2 (щель – внешнее полупространство) получаем систему функциональных уравнений для определения касательных составляющих электрических полей $\vec{e}_{\varsigma_1}$ и $\vec{e}_{\varsigma_2}$ на поверхностях этих щелей:

$$\begin{aligned} &\text{на } \varsigma_1: && \vec{H}_\tau^{in}\left(\vec{e}_{\varsigma_1}\right)+\vec{H}_{0\tau}^{(\mu)}=\vec{H}_\tau^{v}\left(\vec{e}_{\varsigma_1}\right)+\vec{H}_\tau^{v}\left(\vec{e}_{\varsigma_2}\right), \\ &\text{на } \varsigma_2: && \vec{H}_\tau^{v}\left(\vec{e}_{\varsigma_1}\right)+\vec{H}_\tau^{v}\left(\vec{e}_{\varsigma_2}\right)=\vec{H}_\tau^{ext}\left(\vec{e}_{\varsigma_2}\right), \end{aligned} \qquad (1)$$

где $\vec{H}_\tau^{in}\left(\vec{e}_{\varsigma_1}\right)$, $\vec{H}_\tau^{ext}\left(\vec{e}_{\varsigma_2}\right), \vec{H}_\tau^{v}\left(\vec{e}_{\varsigma_1}\right)$, $\vec{H}_\tau^{v}\left(\vec{e}_{\varsigma_2}\right)$ – тангенциальные составляющие магнитного поля, возбуждаемого электрическими полями $\vec{e}_{\varsigma_1}$ и $\vec{e}_{\varsigma_2}$ на соответствующих поверхностях щелей в объемах "*in*", "*ext*", и "*v*", $\vec{H}_{0\tau}^{(\mu)}$ – магнитное поле волны во внутренней области, возбуждающее щель ("μ" – обобщенный индекс, обозначающий тип волны).

Для решения задачи определения тангенциальных составляющих электрического поля на поверхностях щели был применен метод Галеркина с использованием ограниченного количества тригонометрических базисных функций полной области. Электрическое поле на поверхностях ς_1 и ς_2 представлено в виде линейной комбинации координатных базисных функций $\vec{e}_{q1}$ и $\vec{e}_{q2}$:

$$\vec{e}_{\varsigma_1}=\sum_{q=1}^{Q} V_{q1}\vec{e}_{q1},\ \vec{e}_{\varsigma_2}=\sum_{q=1}^{Q} V_{q2}\vec{e}_{q2}. \qquad (2)$$

Так как $L>>d$ и $d<\lambda$, то можно считать, что электрическое поле направлено поперек щели и постоянно в этом направлении. Тогда векторные функции $\vec{e}_{q1}$ и $\vec{e}_{q2}$ имеют вид

$$\vec{e}_{q1}=\vec{e}_{q2}=\vec{y}^{0}\frac{1}{d}\sin\frac{q\pi}{l}z, \qquad (3)$$

где $\vec{y}^{0}$ – орт вдоль оси $\{0y\}$.

Количество и номера собственных функций щели выбираются исходя из соотношений между длиной щели и длинами волн в свободном полупространстве и в волноводе [2]. Задача сводится к решению системы линейных алгебраических уравнений (СЛАУ) относительно искомых коэффициентов в разложении электрического поля на щели V_{q1} и V_{q2}:

$$\begin{cases} \sum\limits_{q=1}^{Q} V_{q1}\left(Y_{pq,11}^{in}+Y_{pq,11}^{v}\right)+\sum\limits_{q=1}^{Q} V_{q2}Y_{pq,12}^{v}=F_{p1}^{(\mu)},\ p=1,2...Q \\ \sum\limits_{q=1}^{Q} V_{q1}Y_{pq,21}^{v}+\sum\limits_{q=1}^{Q} V_{q2}\left(Y_{pq,22}^{v}+Y_{pq,22}^{ext}\right)=0,\ p=1,2...Q \end{cases} \qquad (4)$$

Матричные элементы СЛАУ – частичные проводимости щели $Y_{pq,mn}^{in,v,ext}$ ($m=1,2$, $n=1,2$) определяются следующим образом:

$$Y_{pq,mn}^{in,v,ext}=\int_{\varsigma_m}\left[\vec{e}_{pm},\vec{H}^{in,v,ext}\left(\vec{e}_{qn}\right)\right]\vec{n}^{in,v,ext}d\varsigma, \qquad (5)$$

где $\vec{n}^{in,v,ext}$ – орты нормалей к поверхностям интегрирования, обращенные внутрь соответствующих областей.

Так как векторы $\vec{e}_{q1}$ и $\vec{e}_{q2}$ направлены поперек щели, в выражении для проводимостей (5) будет присутствовать только проекция вектора магнитного поля на ось щели $H_z^{in,v,ext}$. Это поле в каждом из объемов может быть определено с помощью магнитных тензорных функций Грина $\hat{G}_m^{in,v,ext}(\vec{r},\vec{r}')$ [5] для векторного потенциала [6]:

$$H_z^{in,v,ext}(\vec{e}_{qn}) = \frac{\vec{z}^0}{4\pi i\omega\mu_a}\left(\frac{\partial^2}{\partial z^2}+k^2\right)\int\limits_{\varsigma_n}\left[\vec{e}_{qn},\vec{n}^{in,v,ext}\right]\hat{G}_m^{in,v,ext}(\vec{r},\vec{r}')\,d\varsigma,$$

где $\vec{z}^0$ – орт вдоль оси $\{0z\}$.

Правая часть системы уравнений (4) сформирована парциальными магнитодвижущими силами [6]:

$$F_{p1}^{(\mu)} = \int\limits_{\varsigma_1}\left[\vec{e}_{p1}\vec{H}_{0\tau}^{(\mu)}\right]\vec{n}^{ext}d\varsigma.$$

Решение системы уравнений (4) позволяет определить поля $\vec{e}_{\varsigma_1}$ и $\vec{e}_{\varsigma_2}$, а следовательно, и электродинамические характеристики щелевого излучателя: коэффициенты излучения и отражения, диаграмму направленности, коэффициент направленного действия, коэффициент усиления.

В предположении, что волновод возбуждается волной единичной мощности, коэффициент излучения можно найти следующим образом:

$$|S_\Sigma|^2 = \frac{1}{2}\mathrm{Re}\sum_{p=1}^{Q}\sum_{q=1}^{Q}V_{p2}V_{q2}^{*}Y_{pq,22}^{ext},$$

где V_{q2}^{*} – комплексно сопряженная величина.

Диаграмма направленности электрически длинной щели в плоскости вектора $\vec{H}$ с учетом (2), (3) определяется выражением

$$F_H(\theta) = \sin\theta\sum_{q=1}^{Q}V_{q2}\frac{q\pi l\left(1-(-1)^q e^{ikl\cos\theta}\right)}{(q\pi)^2-(kl\cos\theta)^2},$$

где угол θ отсчитывается от продольной оси щели.

Коэффициент усиления волноводно-щелевого излучателя можно определить по формуле $G = D\eta = D|S_\Sigma|^2$, где D – коэффициент направленного действия, η – коэффициент полезного действия, который в случае пренебрежения потерями в стенках волновода совпадает с коэффициентом излучения.

Численные результаты

Как показали исследования, при заполнении полости электрически длинной щели ($L > (20\div30)\lambda$) диэлектриком с небольшим значением диэлектрической проницаемости $\varepsilon^v \approx 1,2\div2,5$ можно значительно понизить уровень боковых лепестков ДН, увеличить КНД и КУ излучателя при возбуждении как основным типом волны, так и высшими модами.

Наличие диэлектрика в полости щели приводит к уменьшению осцилляций в амплитудном распределении поля вдоль щели и, как следствие – к улучшению характеристик направленности.

В качестве примера рассмотрим щель с размерами $l = 1280$ *мм*, $d = 1$ *мм* и $h = 1$ *мм*, прорезанную в середине узкой стенки прямоугольного волновода сечением $a \times b = 23 \times 5$ *мм*^{2} при возбуждении волной типа H_{10} с длиной волны в свободном пространстве $\lambda = 32{,}0$ *мм*, волной типа H_{20} с $\lambda = 17{,}1$ *мм* и волной типа H_{30} с $\lambda = 11{,}56$ *мм*. Длины волн выбраны таким образом, чтобы главные лепестки диаграмм направленности щелевого излучателя на данных длинах волн имели одинаковое направление. Графики на рис. 2 иллюстрируют влияние диэлектрика в полости щели на энергетические и направленные свойства щели.

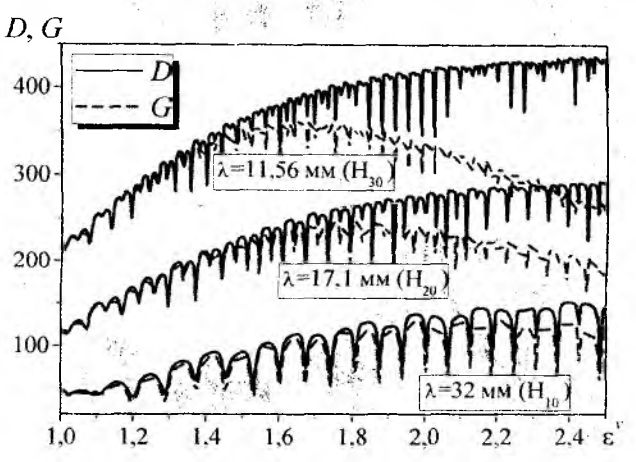


Рис. 2

Как видно из рис. 2, благодаря заполнению полости щели диэлектриком можно значительно увеличить КНД и КУ излучателя (по сравнению со случаем $\varepsilon^v = 1$) при возбуждении как основным типом волны, так и высшими типами волн. С возрастанием значения диэлектрической проницаемости диэлектрика ε^v КНД в среднем увеличивается, стремясь к значению, определяемому как $4l/\lambda$, а КУ при некоторой величине ε^v, зависящей от электрической длины щели, достигает максимального значения и при дальнейшем росте ε^v уменьшается, что связано с уменьшением коэффициента излучения щели.

Значения коэффициентов излучения $|S_\Sigma|^2$, коэффициентов направленного действия D и усиления G щели, полость которой заполнена диэлектриком с параметрами $\varepsilon^v = 2{,}1$, $\operatorname{tg}\delta^v = 10^{-4}$ приведены в таблице, где также представлены значения соответствующих параметров для полой щели ($\varepsilon^v = 1$).

Тип волны	λ, *мм*	$\frac{l}{\lambda}$	$\lvert S_\Sigma \rvert^2$		D		G	
			$\varepsilon^v = 2{,}1$	$\varepsilon^v = 1$	$\varepsilon^v = 2{,}1$	$\varepsilon^v = 1$	$\varepsilon^v = 2{,}1$	$\varepsilon^v = 1$
H_{10}	32,0	40	0,94	0,97	135,8	50,3	128,0	48,7
H_{20}	17,1	74,9	0,82	0,98	279,6	117,7	229,8	117,3
H_{30}	11,56	110,7	0,76	0,99	425,2	210,1	324,4	210,0

Диаграммы направленности щели, заполненной диэлектриком, при возбуждении волнами H_{10}, H_{20} и H_{30} представлены на рис. 3. Введением диэлектрика в полость щели удается уменьшить уровни боковых лепестков и ширину главного лепестка диаграмм направленности. Для сравнения на рисунке приведены ДН полой щели с такими же размерами. Они имеют гораздо более высокие уровни боковых лепестков и более широкие главные максимумы.

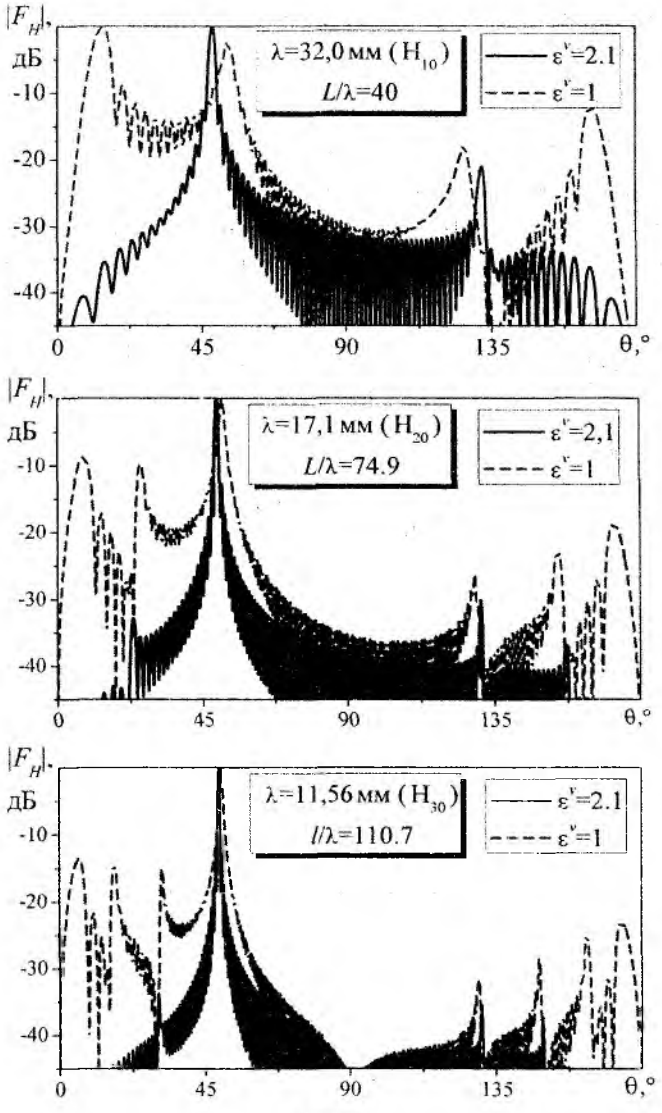


Рис. 3

Таким образом, заполнение полости электрически длинной щели диэлектриком позволяет увеличить коэффициент направленного действия и коэффициент усиления излучателя, прорезанного в узкой стенке прямоугольного волновода, возбуждаемого основным и высшими типами волн. Это дает возможность реализовать на основе такой щели многочастотный излучатель при возбуждении волновода несколькими типами волн на разных частотах.

Список литературы: 1. *Катрич, В.А., Лященко, В.А., Бердник, С.Л.* Электрически длинные волноводно-щелевые антенны с оптимальными излучающими и направленными характеристиками // Известия вузов. Радиоэлектроника. – 2003. – №2. С.51 – 60. 2. *Катрич, В.А. Лященко, В.А., Бердник, С.Л., Пшеничная, С.В.* Оптимизация характеристик излучения электрически длинных продольных щелей в узкой стенке прямоугольного волновода // Радиофизика и радиоастрономия. – 2004. – №4. – С.439-447. 3. *Уолтер, К.Х.* Антенны бегущей волны. – М. : Энергия, 1970. – 447 с. 4. *Lampariello, P., Frezza, F., Shigesawa, H., Tsuji, M., Oliner A.A.* Versatile Leaky-Wave Antenna Based On Stub-Loaded Rectangular Waveguide // IEEE Trans. Antennas and Propagat. – 1998. – Vol. 46, No. 7. – P. 1032 – 1055. 5. *Nesterenko, M.V., Katrich, V.A., Penkin, Y.M., Berdnik, S.L.* Analytical and Hybrid Methods in the Theory of Slot-Hole Coupling of Electrodynamic Volumes. – New York: Springer Science+Business Media, 2008. – 146 p. 6. *Фельд, Я.Н., Бененсон,С.Л.* Антенно-фидерные устройства. – М. : Изд-во ВВИА им. Жуковского, 1959. – Ч. 2. – 551 с.

Харьковский национальный университет
имени В.Н. Каразина

Поступила в редколлегию 20.09.2012

УДК 681.7.068.4

О.І. ФИЛИПЕНКО, д-р техн. наук, О.В. СІЧОВА

МОДЕЛЮВАННЯ ВПЛИВУ СТРУКТУРИ ФОТОННО-КРИСТАЛІЧНИХ ВОЛОКОН НА РОЗПОДІЛ МОДОВОГО ПОЛЯ ТА ВТРАТИ ОПТИЧНОГО СИГНАЛУ В ЇХ З'ЄДНАННЯХ

Введення

Завдяки постійному вдосконаленню технології виготовлення фотонно-кристалічних волокон (ФКВ), значення втрат при розповсюджені світла в ФКВ наближуються до втрат, які властиві стандартним одномодовим волокнам (ОВ) [1, 2]. Можна виділити внутрішні фактори втрат, обумовлені фізико-хімічними явищами в оптичних волокнах, а також викликані відхиленнями конструктивно-технологічних параметрів волокна, і зовнішні, які мають місце при виконанні монтажу. В даній статті розглянуто фактори втрат у з'єднаннях волокон, розроблено декілька моделей ФКВ з різною структурою поперечного перетину, шляхом програмного моделювання отримано їх розподіли поля основної моди, а також вирахувано втрати при оптико-геометричних відхиленнях, що виникають під час з'єднання волокон, та досліджено залежність затухання оптичного сигналу від величини зсувів.

Загальна характеристика проблеми

Під час застосовування та монтажу елементів функціональної електроніки на ФКВ виникає необхідність виконання їх з'єднання одне з одним або, найчастіше, зі стандартним оптичним волокном. Як і у випадку з'єднання стандартних оптичних волокон, з'єднання ФКВ може бути роз'ємним або нероз'ємним. При цьому неминуче виникають зсуви оптичних волокон одне відносно іншого, а при зварюванні волокон, крім того, відбувається руйнування повітряних отворів в оболонці ФКВ та порушується їх порядок розташування, що впливає на розподіл модового поля. На рис.1 представлено фактори, які впливають на втрати оптичного сигналу для випадків роз'ємного та зварного з'єднань ФКВ.

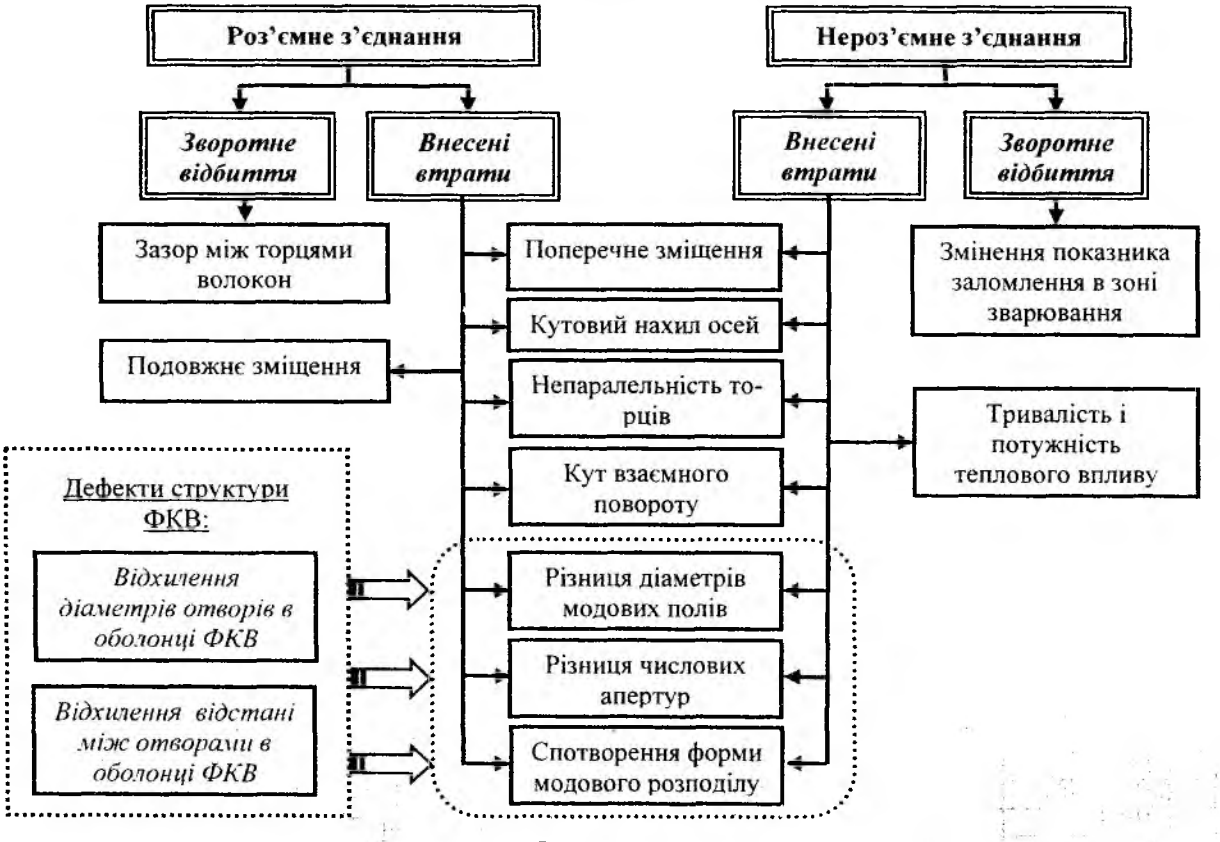


Рис. 1

Найсуттєвішими по внесеним втратам можна виділити поперечне, кутове та подовжн зміщення. Основною вимогою для якісного з'єднання волокон є низькі втрати на стику. Але внаслідок своєї складної періодичної структури поперечного перетину ФКВ цей процес виявляється набагато трудомістким та потребує більшої точності виконання та контролю, ніж у випадку з'єднання стандартних ОВ. Тому важливою задачею являється контроль оптико-геометричних параметрів з'єднань ФКВ під час виконання монтажу.

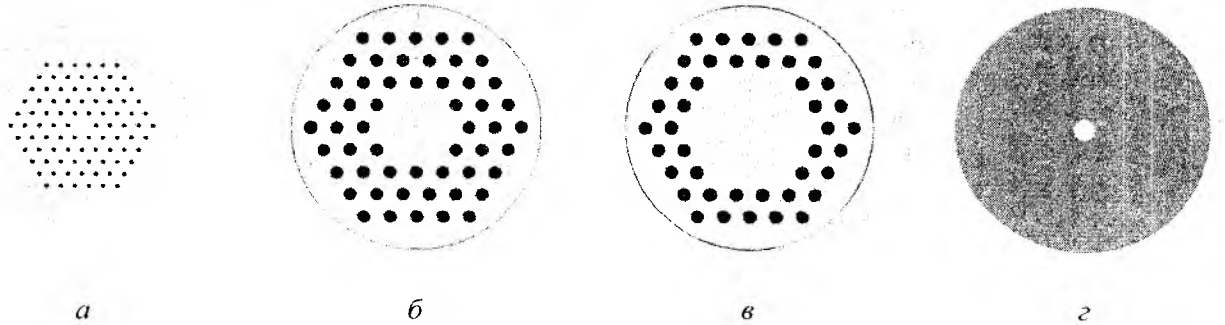
Постановка завдань дослідження

Метою даного дослідження є визначення залежності втрат сигналу від оптико-геометричних відхилень волокон, що з'єднуються. Для досягнення мети необхідно вирішити наступні завдання:

- визначити залежність розподілу модового поля від структури ФКВ (розміру отворів в оболонці, відстані між ними, діаметру серцевини);
- знайти залежність втрат оптичного сигналу від різниці модових полів волокон;
- розрахувати втрати потужності сигналу в з'єднаннях оптичних волокон різного типу при поперечному, подовжньому та кутовому зсуві.

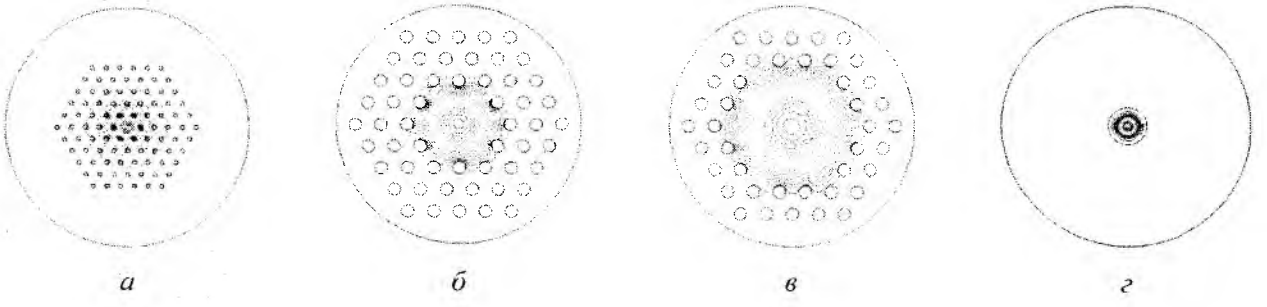
Аналіз впливу структури поперечного перетину ФКВ на розподіл модового поля

Структура ФКВ відрізняється від структури звичайного оптичного волокна. В ОВ є стандартний діаметр серцевини, тоді як діаметр серцевини ФКВ може змінюватися, залежно від його області застосування. На рис. 2 показано змодельовані поперечні перетини ФКВ (*а – в*) та стандартного оптичного волокна з радіусом серцевини $r = 5 мкм$ (*г*). На рис. 3 представлено відповідні до них розподіли поля основної моди на довжині хвилі $\lambda = 1.55 мкм$.



а *б* *в* *г*

Рис. 2



а *б* *в* *г*

Рис. 3

В табл.1 представлено характеристики використаних в моделюванні оптичних волокон.

Таблиця 1

Зображення (позначення)	Діаметр серцевини, *мкм*	Діаметр отворів в оболонці ФКВ, *мкм*	Відстань між отворами, *мкм*	Показник заломлення оболонки	Показник заломлення серцевини	Радіус модового поля, *мкм*
а (PCF-10)	10	2.142	7.14	1.46	-	23
б (PCF-40)	39.4	6.402	13.2	1.4457	-	32.5
в (PCF-60)	63.0	6.402	13.2	1.4457	-	48.5
г (SMF)	10	-	-	1.4378	1.4457	12.5

Втрати оптичного сигналу при з'єднані ФКВ з різними розподілами модового поля

Показником якості з'єднання являється коефіцієнт передачі оптичного сигналу з одного волокна в інше. В випадках відсутності зміщень в з'єднані, він може бути визначений приблизно таким чином [3, 4]:

$$T \approx \frac{4A_{eff1} \cdot A_{eff2}}{\left(A_{eff1} + A_{eff2}\right)^2}, \tag{1}$$

де A_{eff1}, A_{eff2} – ефективні площі модових полів волокон, що з'єднуються:

$$A_{eff} = \frac{\left(\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} |E(x,y)|^2 \, dxdy\right)^2}{\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} |E(x,y)|^4 \, dxdy}, \tag{2}$$

де $E(x,y)$ – амплітуда електричного поля.

Втрати в цьому випадку можна визначити по формулі

$$\alpha = -10\lg(T). \tag{3}$$

Якщо з'єднуються два ідентичних волокна з однаковим розподілом модового поля, коефіцієнт передачі в ідеальному випадку буде дорівнювати одиниці, а втрати, відповідно, нулю. Однак на практиці такі значення недосяжні внаслідок різних факторів, які впливають на якість з'єднання (див. рис. 1).

Визначення втрат сигналу в з'єднаннях ФКВ з урахуванням величини зсувів

Втрати при поперечному зсуві волокон можна вирахувати по формулі

$$\alpha_t = -10\lg\left(\frac{4w_1^2 w_2^2}{\left(w_1^2 + w_2^2\right)^2} \exp\left(-\frac{2r_d^2}{w_1^2 + w_2^2}\right)\right), \tag{4}$$

де $r_d = \left(d_x^2 + d_y^2\right)^{1/2}$ – поперечне зміщення двох волокон; w_1, w_2 – їх радіуси модових полів.

Втрати при подовжньому зміщенні волокон можна визначити

$$\alpha_t = -10\lg\left(\frac{4w_1^2 w_2^2}{\frac{l^2}{k_0^2} + \left(w_1^2 + w_2^2\right)^2}\right), \tag{5}$$

де l – подовжнє зміщення волокон; $k_0 = 2\pi/\lambda$.

Втрати при кутовому зміщенні визначаються по формулі

$$\alpha_\theta = -10\lg\left(\frac{4w_1^2 w_2^2}{\left(w_1^2 + w_2^2\right)^2} \exp\left(-\frac{k^2\theta^2 w_1^2 w_2^2}{2\left(w_1^2 + w_2^2\right)}\right)\right), \tag{6}$$

де θ – кутове зміщення волокон [5].

Результати розрахунків по формулам (4) – (6) для різних типів з'єднань оптичних волокон представлено в табл.2.

За результатами розрахунків було побудовано графіки залежностей втрат оптичного сигналу від поперечного зсуву (рис. 4), подовжнього (рис. 5) та кутового (рис. 6) зміщення.

Таблиця 2

№	Тип з'єднання	r_d, *мкм*	α_t, *дБ*	l, *мкм*	α_l, *дБ*	θ, *град.*	α_θ, *дБ*
1	PCF-10/SMF	0.5	3.516	0.5	3.508	0.2	3.513
		2.5	3.691	2.5	3.508	0.6	3.551
		4.5	4.099	4.5	3.508	0.8	3.585
2	PCF-10/PCF-10	0.5	0.005	0.5	$0.014 \cdot 10^{-3}$	0.4	0.012
		2.5	0.118	2.5	$0.340 \cdot 10^{-3}$	0.6	0.028
		4.5	0.383	4.5	$1.101 \cdot 10^{-3}$	0.8	0.050
3	PCF-40/PCF-40	0.5	0.005	0.5	$0.034 \cdot 10^{-6}$	0.4	0.006
		2.5	0.059	2.5	$0.852 \cdot 10^{-6}$	0.6	0.014
		4.5	0.192	4.5	$2.761 \cdot 10^{-6}$	0.8	0.025
4	PCF-60/PCF-60	0.5	0.001	0.5	$0.007 \cdot 10^{-6}$	0.4	0.003
		2.5	0.027	2.5	$0.172 \cdot 10^{-6}$	0.6	0.006
		4.5	0.086	4.5	$0.557 \cdot 10^{-6}$	0.8	0.011
5	PCF-40/PCF-60	0.5	1.563	0.5	1.562	0.4	1.565
		2.5	1.598	2.5	1.562	0.6	1.570
		4.5	1.680	4.5	1.562	0.8	1.577
6	PCF-40/SMF	0.5	8.011	0.5	8.007	0.4	8.017
		2.5	8.110	2.5	8.007	0.6	8.031
		4.5	8.340	4.5	8.007	0.8	8.050
7	PCF-60/SMF	0.5	14.540	0.5	14.540	0.4	14.55
		2.5	14.590	2.5	14.540	0.6	14.55
		4.5	14.700	4.5	14.540	0.8	14.56

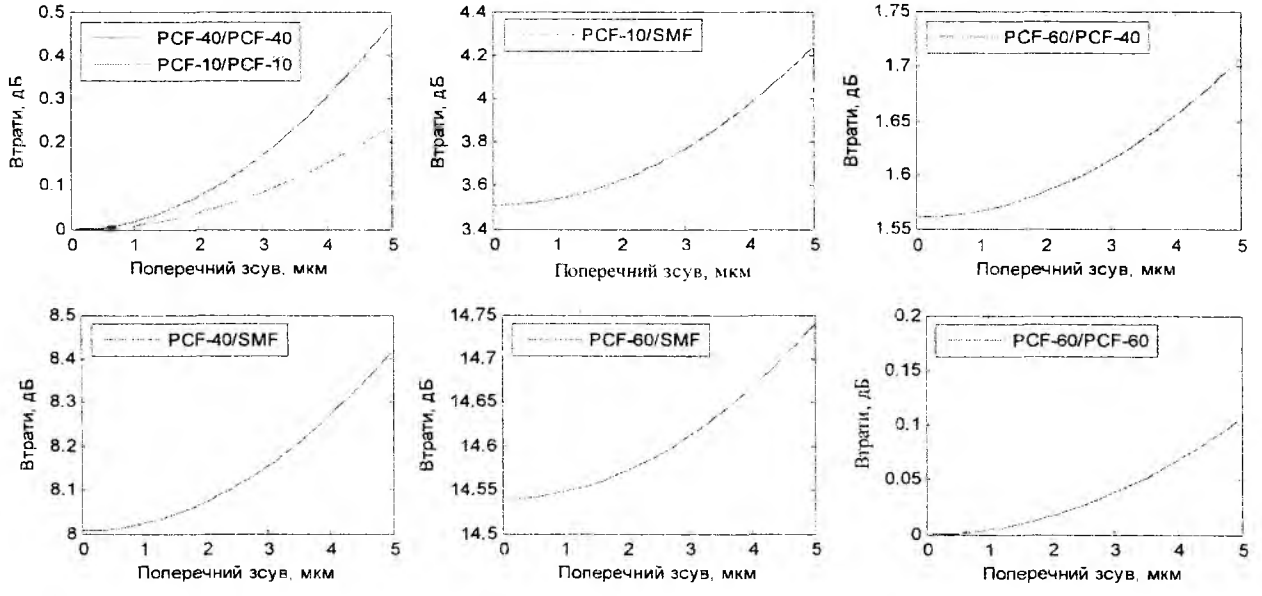


Рис. 4

Проаналізувавши отримані результати проведених досліджень можна заключити, що при поперечному зміщенні двох ідентичних ФКВ втрати будуть більші для волокон з меншим діаметром модового поля, але не досягають 0.5 *дБ* при зсувах до 5 *мкм*. Тоді як на з'єднаннях ФКВ та стандартного ОВ або двох ФКВ з різними розмірами модових полів втрати збільшуються та досягають значень до 14.5 *дБ* (для випадку з'єднання ФКВ з радіусом модового поля 48.5 *мкм* та стандартного ОВ). Втрати при подовжньому зміщенні двох ФКВ з однаковою структурою ще більш незначущі, а при з'єднані волокон з різними діаметрами модового поля затухання приблизно таке, як і для поперечного зсуву – до 14.5 *дБ* для величини зазору до 5 *мкм*. Така ж сама картина спостерігається і для кутового зміщення.

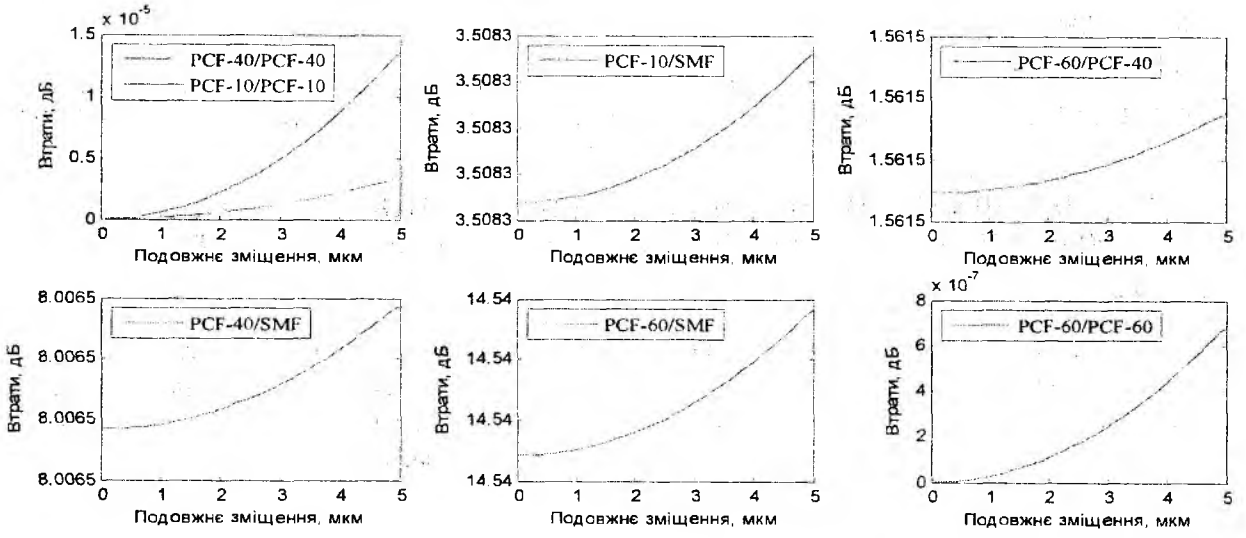


Рис. 5

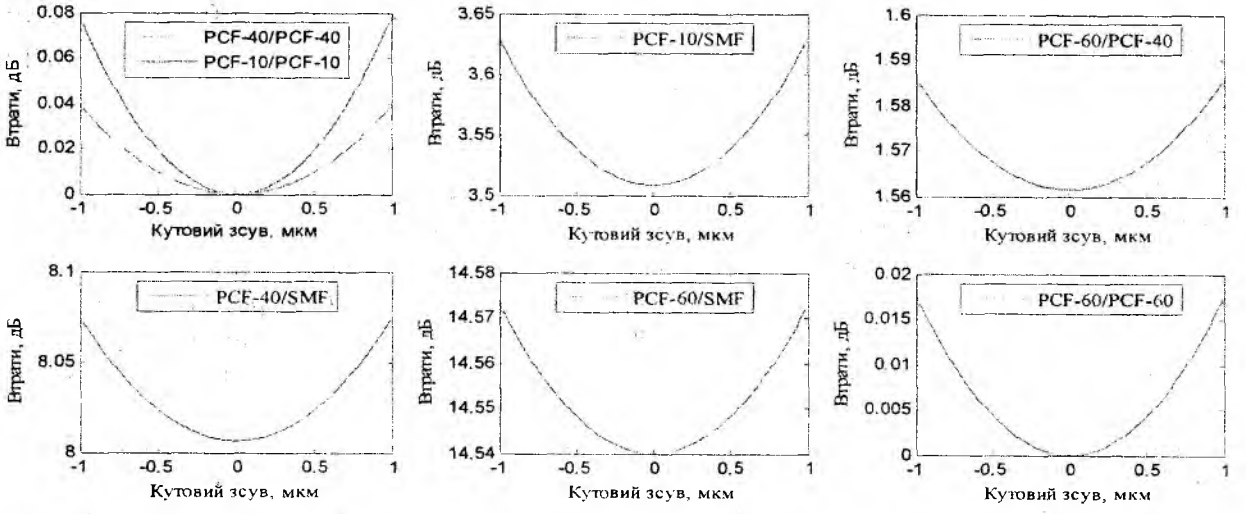


Рис. 6

Висновки

В статті проаналізовано фактори втрат оптичного сигналу в з'єднаннях ФКВ, зазначено, що найбільший вплив на затухання мають поперечне, подовжнє та кутове зміщення. За допомогою комп'ютерного моделювання побудовано поперечні перетини оптичних волокон чотирьох різних типів та отримано їх розподіли полів основної моди. З отриманих зображень видно, що структура волокна (розмір отворів в оболонці, відстань між ними, діаметр серцевини) впливає на розподіл модового поля. Для декількох варіантів з'єднань оптичних волокон розраховано втрати потужності сигналу та побудовано графіки залежності затухання від зміщень з урахуванням розмірів модових полів волокон. Подальші дослідження в цій області направлені на вивчення впливу інших оптико-геометричних параметрів з'єднань ФКВ на передачу сигналу (таких як кут взаємного повороту, непаралельність та шорсткість торців волокон, різниця числових апертур).

Список літератури: 1. *Naniy, O. E., Pavlova, E. G.* Photonic crystal fibers // Lightwave Russian Edition. – 2004. – №3. – P.47-53. 2. *Pavlova, E. G.* Mechanisms of losses in photonic crystal fibers // Lightwave Russian Edition. – 2005. – №3. – P.54-56. 3. *W. Zhi , R. Guobin, L. Shuqin, J. Shuisheng.* Supercell lattice method for photonic crystal fibers // OPTICS EXPRESS, 5 May 2003, Vol.11, No9, pp.980-991. 4. *Mortensen, N. A.* Effective area of photonic crystal fibers // OPTICS EXPRESS, 8 April 2002, Vol.10, No7, pp.341-348. 5. *Kliros, G.S. Konstantantinidis, J., Thraskias, C.* Prediction of Macrobending and splice losses for photonic crystal fibers based on the effective index method // Wseas transactions on communications. – 2006. – Vol.5, Issue 8. – P.1314-1321.

Харківський національний
університет радіоелектроніки

Надійшла до редколегії 15.09.2012

УДК 621.315.592

Н.И. СЛИПЧЕНКО, д-р. физ.-мат. наук, В.А. ПИСЬМЕНЕЦКИЙ, канд. техн. наук, А.В.ФРОЛОВ, канд. техн. наук, Н.В. ГЕРАСИМЕНКО, М.Ю. ГУРТОВОЙ, Е.С. ГЛУШКО, Т.Е. СТЫЦЕНКО

ИССЛЕДОВАНИЕ И ОПТИМИЗАЦИЯ ПЛЕНОЧНЫХ КРЕМНИЕВЫХ АМОРФНЫХ ФОТОПРЕОБРАЗОВАТЕЛЕЙ НА *p-i-n*-СТРУКТУРАХ

Введение

В последнее время при изготовлении пленочных фотопреобразователей аморфный кремний выступил в качестве более дешевой альтернативы монокристаллическому благодаря гораздо большему оптическому поглощению (в 20 раз), малой толщине пленок a-Si:H (0,5 – 1,0 *мкм*) и, как следствие, сокращению расхода кремния. Кроме этого, достоинством пленочной технологии являются более низкие рабочие температуры ($T \leq 300^{O}C$).

Более высокой эффективностью преобразования обладают солнечные элементы на основе аморфного кремния с *p-i-n*-структурой [1]. Для ее эффективной работы необходимо обеспечить поглощение основной части светового потока, генерацию фотоносителей заряда и их ускорение в *i*-области под действием электрического поля.

Основная часть

В процессе выполнения работы планировалось проведение исследований:

- солнечной ячейки на основе пленочной кремниевой *p-i-n*-структуры;
- тандемного пленочного фотопреобразователя на двух *p-i-n*-ячейках.

Кроме этого было предусмотренно проведение оптимизации параметров верхней и нижней *p-i-n*-ячеек тандемных пленочных фотопреобразователей и выполнение экспериментальных исследований подсистемы солнечной станции, состоящей из аморфных кремниевых панелей BS-50.

Оценка исходных параметров *p-i-n*-фотопреобразователей (ФП) была выполнена на основе оптических характеристик солнечного излучения при потоке 100 $мВт/см^2$.

Для решения этой задачи были использованы известные соотношения для спектрального энергетического S(ε) и удельного спектрального потока фотонов $S_{уд}(\varepsilon)$ [3].

Для вычисления плотности тока генерации *p-n*-перехода интегрируем спектральное распределение $S_{уд}(\varepsilon)$, а для перехода к фототоку вводим коэффициент собирания K_{SO}, учитывающий различные виды потерь фотогенерированых носителей заряда. Тогда согласно [3] для плотности тока генерации $J_Г$ фототока верхнего *p-n*-перехода $J_{ф1}$ и фототока нижнего *p-n*-перехода $J_{ф2}$ можно применить расчетные формулы:

$$j_Г = K_{SO} \cdot e \cdot \int_{\varepsilon}^{\infty} S_{уд}(\varepsilon) d\varepsilon \; ; \tag{1}$$

$$j_{Ф1} = K_{SO} \cdot e \cdot \int_{\varepsilon 1}^{\infty} S_{уд}(\varepsilon) d\varepsilon \; ; \tag{2}$$

$$j_{Ф2} = K_{SO} \cdot e \cdot \int_{\varepsilon 1}^{\varepsilon 2} S_{уд}(\varepsilon) d\varepsilon \; . \tag{3}$$

Для определения напряжения холостого хода используем известное соотношение [2] при $J_{кз} \approx J_ф$:

$$U_{XX} = U_t \cdot \ln\left(\frac{J_Ф}{J_S} + 1\right),$$

где U_t – температурный потенциал, равный U_t =26 *мВ* при температуре 300 К.

Поскольку рекомбинационный ток возрастает с увеличением напряжения гораздо медленнее ($\sim e^{eU/2kT}$), чем обратный ток, связанный с током насыщения ($\sim e^{eU/kT}$) при высокой плотности излучения роль рекомбинационного тока меньше.

Для улучшения собирания необходимо расширение *i*-слоя, ширина которого должна считаться сопоставимой с диффузионной длиной L_n неосновных носителей (электронов) в базе. При расчетах полагаем, что интенсивность падающего излучения довольно высока и концентрация неравновесных носителей в *i*-области существенно превышает концентрацию равновесных.

Таким образом, считая, что на поверхность фотоэлемента падает поток монохроматических фотонов S(ε), ток короткого замыкания имеет вид [2]:

$$i_{\Phi} = ei_{\Phi} = eS(\varepsilon)(1-e^{-\alpha d}) + eS(\varepsilon)\cdot\frac{e^{-\alpha d}\cdot\alpha L_n}{1+\alpha L_n},$$

где α – коэффициент поглощения.

Для определения ВАХ фотоэлемента рассмотрим концентрацию неравновесных носителей в *i*-слое, ширина которого мала по сравнению с параметрами L_n. Поэтому концентрация неравновесных носителей постоянна во всех точках *i*-слоя. Концентрация электронов равна концентрации дырок, так как в противном случае должны были бы возникнуть аномально сильные электрические поля. При $n \approx p >> n_i$ число рекомбинирующих носителей в *i*-слое в единицу времени:

$$\frac{nd}{\tau} = \frac{pd}{\tau},$$

где $\tau = \tau_n + \tau_p$.

Если при выбранной ширине слоя *d* решающую роль играет рекомбинационный ток, то:

$$i_{\phi} - i = \frac{epd}{\tau};$$

$$p = n = \frac{i_{\Phi} - i}{d\tau}, \qquad (4)$$

где *i* – ток нагрузки.

Для расчета выходного напряжения фотоэлемента используем выражения для контактной разности потенциалов на *n-i* и *i-p*-переходах [2]. При этом общее изменение контактной разности потенциалов между *n* и *p* областями с промежуточным *i*-слоем имеет вид

$$\varphi_k = \frac{kT}{e}\cdot\ln\frac{n}{n_i}.$$

Если из последней величины вычесть омическое падение напряжения при прохождении тока через *i*-слой, получим выходное напряжение:

$$U = 2\frac{kT}{e}\cdot\ln(\frac{p}{n_i}) - \frac{id}{(\mu_n+\mu_p)\cdot p\cdot e}, \qquad (5)$$

где μ_n и μ_p – подвижности электронов и дырок; $d/[(\mu_n+\mu_p)pe]$ – сопротивление 1 $см^2$ *i*-слоя.

Из (5) с учетом (4) имеем:

$$U = 2\frac{kT}{e}\cdot\ln(\frac{i_{\Phi}-i}{i_r}) - \frac{d^2 i}{(\mu_n+\mu_p)\cdot(i_{\Phi}-i)\cdot\tau}, \qquad (6)$$

где $i_r = en_i d/\tau$ – поток рекомбинации (или генерации) в равновесном состоянии.

Формула (6) дает окончательное выражение для вольтамперной характеристики фотоэлемента с *p-i-n*-структурой.

Освоение нанотехнологий открывает перспективы создания наноразмерных *p-i-n*-диодов (рис. 1). Для расчета их параметров можно воспользоваться раннее приведенными соотношениями.

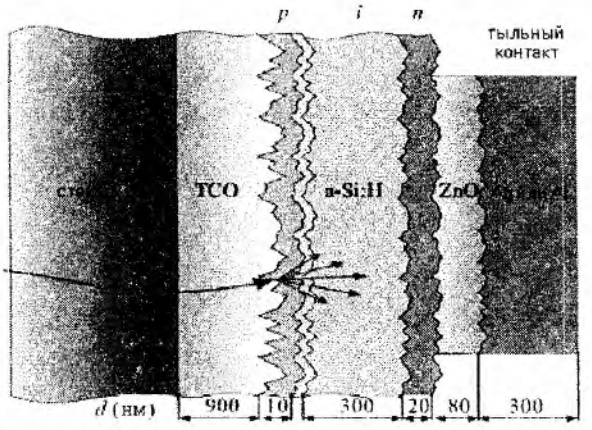


Рис. 1. Структура тонкопленочного однопереходного фотопреобразователя на основе *p-i-n*-ячейки

Результаты расчетов приведены на рис. 2 – 6. На рис. 2 демонстрируется зависимость оптимального выходного напряжения от ширины запрещенной зоны, когда толщина *i*-области принимает значение 0,2; 0,5; 1 и 2 *мкм*.

Вполне понятно, что оптимальное напряжение растет пропорционально увеличению ширины запрещенной зоны. При расширении *i*-области потери напряжения на ее последовательном сопротивлении растут, и соответственно уменьшается оптимальное выходное напряжение.

При росте толщины *i*-области увеличивается количество фотоносителей заряда, что приводит к росту плотности тока и иллюстрируется на рис. 2, *б*.

Уменьшение плотности оптимального тока при росте ширины запрещенной зоны хорошо согласуется со спектральной характеристикой потока фотонов и интегральным соотношением, с помощью которого рассчитывается фототок.

Зависимости напряжения холостого хода от ширины запрещенной зоны фактически повторяют аналогичные зависимости оптимального напряжения на рис.2, *а* при различных значениях толщины *i*-области.

Зависимости выходной удельной мощности от ширины запрещенной зоны вполне понятны как результат перемножения оптимального напряжения на оптимальную плотность тока.

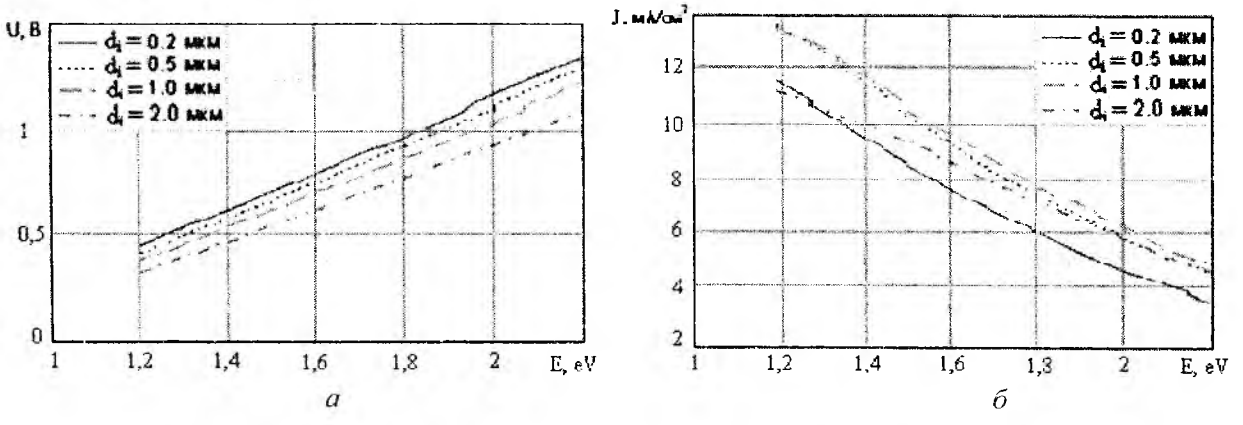
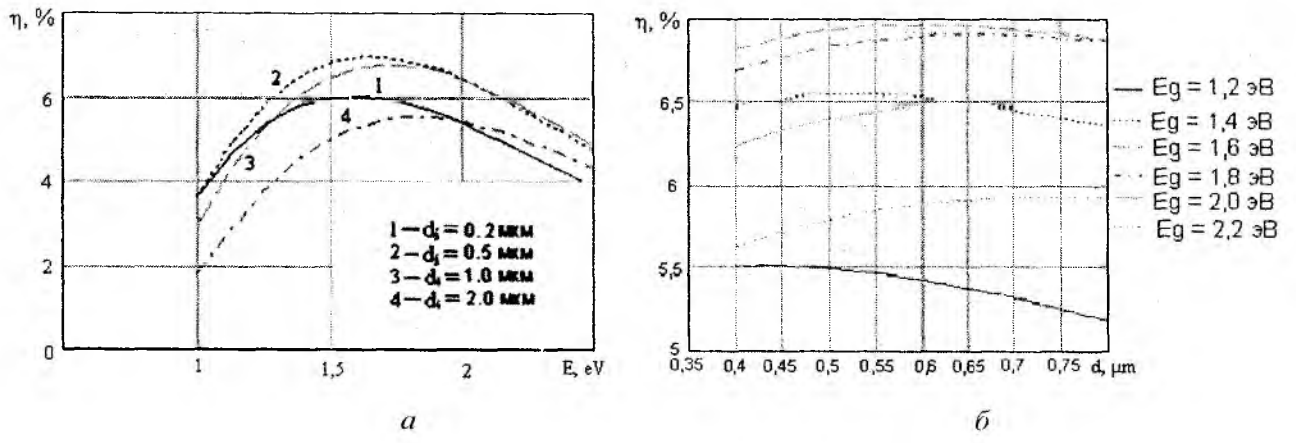


Рис. 2. Зависимость: *а* – оптимального напряжения; *б* -оптимальной плотности тока от ширины запрещенной зоны

Эти зависимости позволяют определить оптимальные значения ширины запрещенной зоны.

Зависимости КПД от ширины запрещенной зоны фактически повторяют аналогичные зависимости удельной мощности.

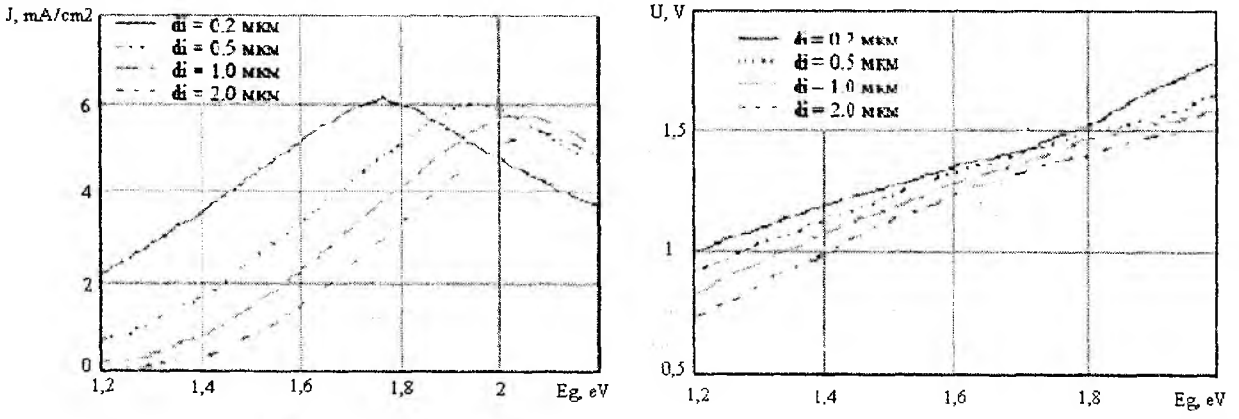


а *б*

Рис.3. Зависимость КПД от ширины: *а* – запрещенной зоны; *б* – *i*-области

На рис.3, *б* приведены зависимости КПД от ширины *i*-области при различных значениях ширины запрещенной зоны. Как видим, существует оптимальная ширина *i*-области, которая полностью совпадает с ранее полученными результатами по оптимальной толщине аморфной пленки в гетероструктурных ФП на основе α-Si:H/Si(c)

На рис. 4 представлены зависимости оптимальной плотности тока и оптимального напряжения соответственно от ширины запрещенной зоны верхней *p-i-n*-ячейки двухпереходного фотопреобразователя. Наличие максимума связано со спектром поглощения пленки ά-Si:H. Смещение максимума обусловлено перемещением спектра поглощения пленки ά-Si:H с уменьшением ее толщины. С ростом толщины *i*-слоя наблюдается рост фототока и падение напряжения на *i*-области. Выходное напряжение растет пропорционально ширине запрещенной зоны.



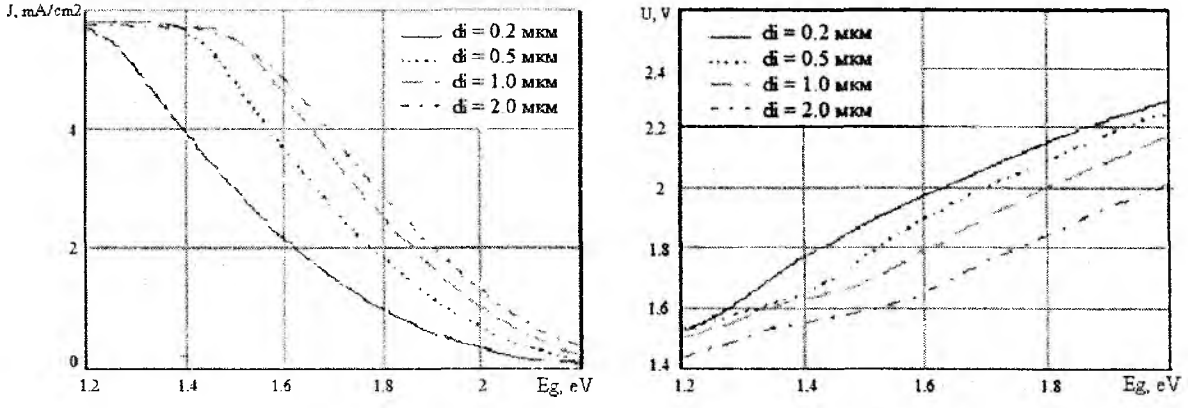
а *б*

Рис. 4. Зависимость: *а* – оптимальной плотности тока; *б* – оптимального напряжения от ширины запрещенной зоны верхней *p-i-n*-ячейки двухпереходного фотопреобразователя

На рис. 4 представлены зависимости оптимальной плотности тока и оптимального напряжения соответственно от ширины запрещенной зоны нижней *p-i-n*-ячейки двухпереходного фотопреобразователя. Уменьшение фототока с ростом Eg связано с уменьшением ширины спектра светового потока. С уменьшением толщины пленки *di* ее спектр поглощения смещается вправо. С ростом параметра *di* увеличивается фототок и падает напряжение. При этом напряжение пропорционально ширине запрещенной зоны.

На рис. 5 приведены зависимости КПД тандемной гетероструктуры с учетом ширины запрещенной зоны E_g для верхней и нижней *p-i-n*-ячеек при разной толщине *i*-слоя. На графиках четко просматривается оптимальное значение параметра E_g на величину которого

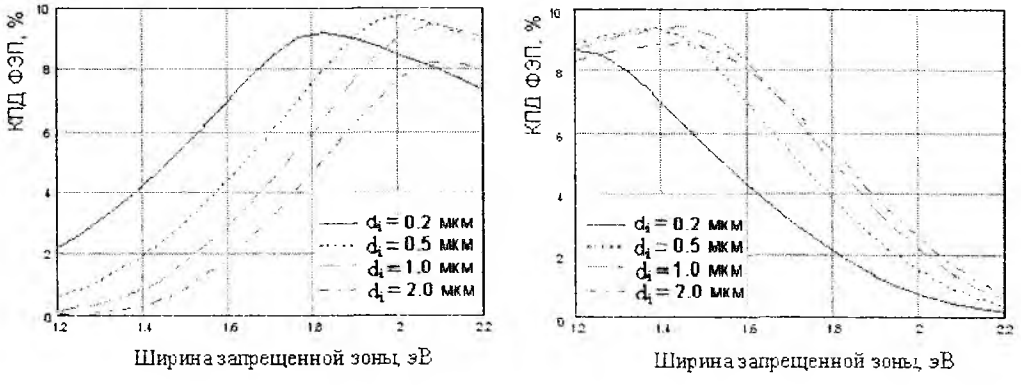
влияет толщина *i*-слоя. Таким образом, целесообразно проведение процедуры оптимизации по значениям параметра E_g верхней и нижней *p-i-n*-ячеек.



а *б*

Рис. 5. Зависимость: *а* – оптимальной плотности тока; *б* – оптимального напряжения от ширины запрещенной зоны нижней *p-i-n*-ячейки двухпереходного фотопреобразователя

Результаты выполнения процедуры оптимизации по максимальному значению КПД с использованием метода градиентного спуска при толщине *i*-слоя, равного 0,5 *мкм*, представлены на рис.7. Оптимальные значения ширины запрещенной зоны E_g составили для верхней *p-i-n*-ячейки 2,075 *эВ* и для нижней – 1,45 *эВ*. Соответственно максимальное значение КПД равно 9,3 %.



а *б*

Рис. 6. Зависимость КПД для: *а* – верхней; *б* – нижней *p-i-n* ячеек тандемного фотопреобразователя

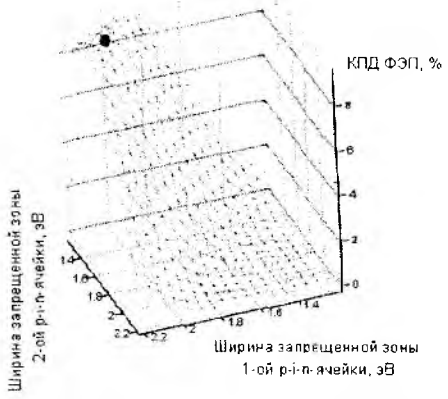
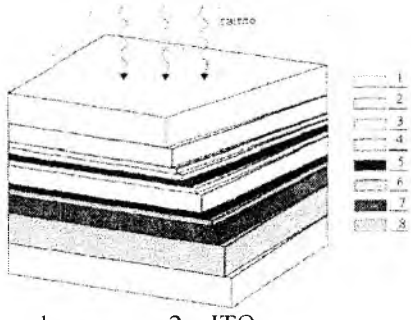


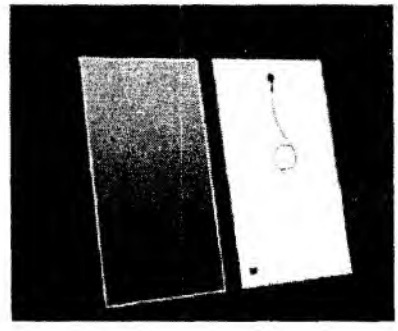
Рис. 7. Поиск оптимальных значений ширины запрещенной зоны обеих *p-i-n*-ячеек тандемной структуры

На рис. 8 приведена структура тандемного ФП, сформированного на двух последовательно соединенных *p-i-n*-переходах с использованием пленок аморфного гидрогенезированного кремния a-Si:H. Такие тандемные ФП изготовляются фирмой BudaSolar, а на их основе солнечные аморфные модули BS-50 с выходной мощностью 50 *Вт*, КПД – 6,3 %, и площадью 0,79 $м^2$.



1 – стекло; 2 – ITO;
3 – пленка p-a-Si:H;
4 – пленка i-a-Si:H;
5 – пленка n-a-Si:H;
6 – слой ZnO;
7 – слой алюминия;
8 – фольга EVA

а



б

Рис. 8. *а* – структура тонкопленочного тандемного двухпереходного *p-i-n* ФП, *б* – солнечный модуль из аморфного кремния BS-50.

Подсистема из модулей модели BS-50 с активной площадью 876 $м^2$ генерирует выходную мощность 54 *кВт* в составе солнечной электростанции мощностью 100 *кВт* для автономного энергообеспечения районного сельскохозяйственного комплекса в Белгородской области (Россия) (рис.9).

Авторами работы (рис. 12) в период с 26 по 28 октября 2011 г. были проведены исследования временных характеристик энергетической подсистемы на основе аморфных модулей BS-50.

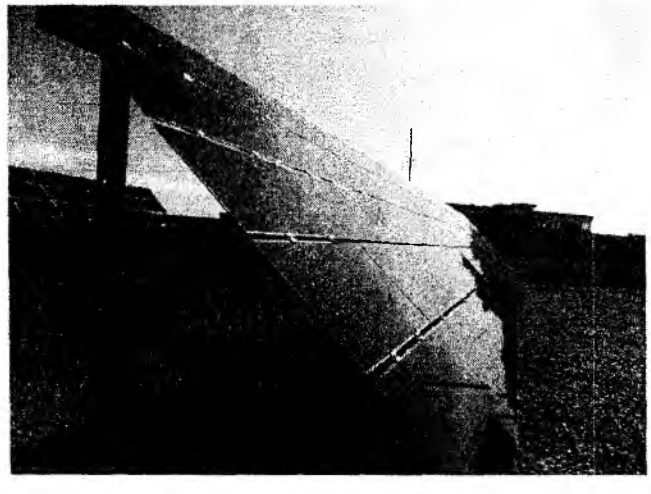


Рис. 9. Подсистема солнечной станции из панелей BS-50

На рис. 10 представлены зависимости освещенности и тока короткого замыкания от времени суток. Из графиков можно сделать вывод, что зависимость тока $I_{кз}$ практически совпадает с графиком изменения освещённости. Это объясняется тем, что для кремниевых монокристаллических преобразователей ток $I_{кз}$ практически совпадает с током $I_ф$, который равен [4, 5]

$$I_{КЗ} = I_{Ф} = \frac{e \cdot \Phi_0}{h\nu} \cdot (1 - R) \cdot \eta \cdot K_C \cdot S ,$$

где Φ_0 – входной световой поток; R – коэффициент отражения светового потока от фронтальной поверхности; η – квантовый выход; K_c – коэффициент собирания фотоносителей; S – площадь фронтальной поверхности.

Из формулы для $I_ф$ следует, что все сомножители кроме параметра Φ_0 являются константами, и поэтому временная зависимость тока $I_{кз}$ действительно воспроизводит график изменения освещенности.

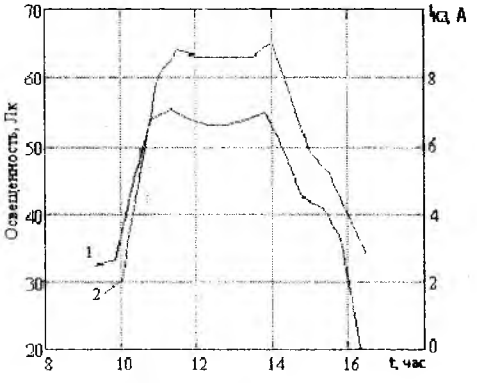


Рис. 10. Зависимости освещенности (1) и тока короткого замыкания (2) от времени суток

Зависимости U xx и Up от времени суток и освещенности представлены на рис. 11.

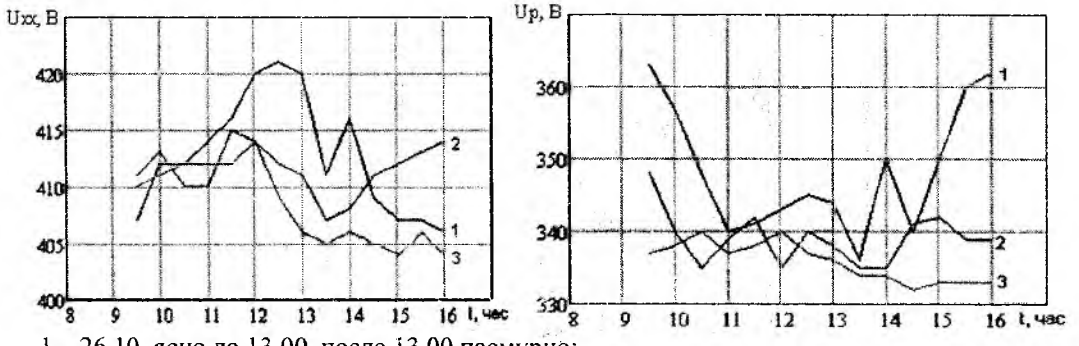
Характерно, что напряжение U_{xx} на этом же отрезке времени практически мало изменяется, что легко объясняется с помощью известного соотношения:

$$\varphi_T = \frac{kT}{e} \qquad I_{КЗ} = \phi_T \cdot \ln\left(\frac{I_Ф}{I_0} + 1\right) ,$$

где I_0 – ток насыщения, *e* – заряд электрона.

В пределах светового дня исследовались временные зависимости напряжения холостого хода U_{xx} (рис. 11, *а*), рабочего напряжения (рис. 11, *б*).

Поскольку температура медленно изменяется с течением времени множитель φ_T в процессе эксперимента практически остается постоянным, как и ток насыщения I_0. Следовательно, параметр U_{xx} связан с током $I_{кз}$ логарифмической зависимостью и поэтому мало изменяется.



1 – 26.10, ясно до 13.00, после 13.00 пасмурно;
2 – 27.10, ясно до 11.00, после 11.00 облачно;
3 – 28.10, пасмурно.

а *б*

Рис. 11. Зависимости: *а* – U xx; *б* – Up от времени суток и освещенности

Проведенные исследования позволяют сделать следующие выводы:

1. Показано что, базовая *p-i-n*-фотоячейка характеризуется оптимальными значениями параметров Eg и di по критерию максимального КПД фотопреобразования.

2. Для нормальной работы тандемной *p-i-*n структуры необходимо сопряжение по току верхней и нижней фотоячеек, которое реализуется с помощью предложенного алгоритма.

3. Оптимизацию по параметру Eg ячеек тандемной *p-i-n* структуры целесообразно выполнять с применением метода градиентного спуска.

4. Проведенные экспериментальные исследования солнечных *p-i-n* панелей типа BS-50 подтвердили целесообразность оптимизации параметров фотоячеек по максимальному КПД.

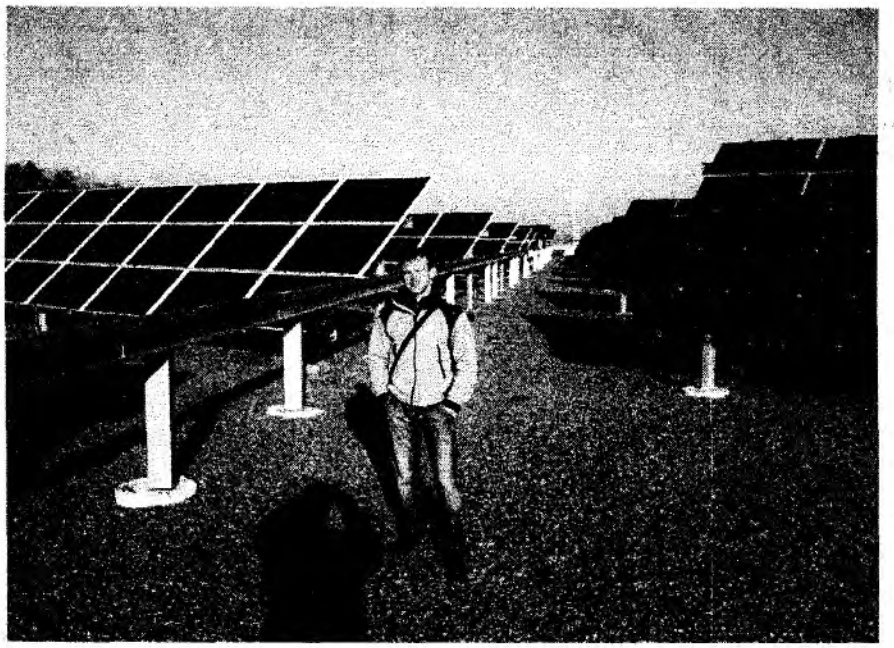


Рис. 12. Общий вид солнечной электростанции
на хуторе Крапивенские Дворы Белгородской области

Авторы выражают искреннюю признательность Генеральному директору ООО «ГК-Агро-Белогорье» В.Ф. Зотову, Генеральному директору ООО «Альт-Энерго» Н.С. Прокопенко за предоставленную возможность проведения исследований выходных фотоэлектрических характеристик солнечной электростанции на хуторе Крапивенские Дворы Белгородской области.

Список литературы: 1. *Мейтин.* Фотовольтаика: материалы, технологии, перспективы. Пусть всегда будет солнце! // ЭЛЕКТРОНИКА: Наука, Технология, Бизнес. – 2000. – №6. – С. 40-46. 2. *Васильев, А.М., Ландсма, А.П.* Полупроводниковые фотопреобразователи. – М. : Сов. радио, 1971. – С.248. 3. *Слипченко, Н.И., Письменецкий, В.А. и др.* Оптимизация параметра состава *x* по выходным характеристикам фотопреобразователей структурой $Al_xGa_{1-x}As$-$In_xGa_{1-x}As$-GaAs // Вестник нац. техн. ун-та «ХПИ». – 2011. – № 54. – С. 144-152. 4. *Бродски, М., Карлсон, Д., Коннел, Дж. и др.* Аморфные полупроводники. – М. : Мир, 1962. – С.. 419. 5. *Слипченко, Н.И., Письменецкий, В.А. и др.*. Экспериментальные исследования характеристик солнечной панели KV 150/24(12) в зимних условиях // Вестник нац. техн. ун-та «ХПИ». – 2011. – №24. – С. 86-92.

Харьковский национальный
университет радиоэлектроники

Поступила в редколлегию 03.09.2012

УДК 528::629.783

А.А. ЖАЛИЛО, канд. техн. наук

РАЗРАБОТКА И ТЕСТИРОВАНИЕ НОВЫХ ЭФФЕКТИВНЫХ МЕТОДОВ И АЛГОРИТМОВ ОБНАРУЖЕНИЯ И УСТРАНЕНИЯ ФАЗОВЫХ СКАЧКОВ СТАТИЧЕСКИХ И КИНЕМАТИЧЕСКИХ ГНСС-НАБЛЮДЕНИЙ

Введение

Точное определение местоположения (позиционирование) с использованием глобальных навигационных спутниковых систем (ГНСС) широко применяется в таких приложениях, как геодезия, картография, межевание земли, кадастр, аэрофотосъемка, ГИС, точное земледелие, геодинамика и др. Все эти приложения предъявляют жесткие требования к точности определения координат – на сантиметровом и даже на миллиметровом уровне. Решение задач высокоточного позиционирования (определения местоположения) с указанной точностью невозможно без использования фазовых наблюдений [1 – 4]. Обработка фазовых наблюдений имеет ряд важных особенностей, которые заключаются в необходимости надежного устранения циклических фазовых скачков (т.н. "слипов") и разрешения фазовой неоднозначности (РФН) [1, 2, 5].

В настоящей работе описываются результаты многолетних разработок, исследований и тестирования (экспериментальной верификации) новых эффективных методов и алгоритмов решения задач определения потерь счета циклов в одночастотных/двухчастотных фазовых ГНСС–наблюдениях и их исправления на этапе послесеансной предварительной обработки и анализа наблюдений.

Сложность проблемы восстановления непрерывности фазовых наблюдений подтверждается постоянным поиском новых более эффективных и более надежных методов и приемов обработки. Существует достаточно обширный список научной и технической литературы, посвященной исследованию этой проблемы. Так, по данной проблеме могут быть выделены работы [6 – 9] и ряд других. В этих работах предложены не только методы решения задачи применительно к конкретным условиям, но также и сравнение характеристик и особенностей других методов, уже известных ранее. Можно рассмотреть основные известные подходы и методы определения (обнаружение, идентификация и оценка) и устранение (исправление) фазовых скачков, которые широко используются в практике обработки ГНСС–наблюдений. Большинство подходов подразумевает формирование и последовательную обработку (с многочисленными разновидностями) линейных комбинаций (ЛК) наблюдений, так или иначе чувствительных к появлению скачков. Далее разработчики и исследователи применяют различные сглаживающие и "дифференцирующие" функции для этих ЛК для целей обнаружения и оценки скачков. Эффективность используемых в настоящее время методов обработки зависит от характеристик наблюдений, вида и частоты следования фазовых скачков, количества пропусков наблюдений, темпа данных, уровня шумов и др.

Как показано в работах [10 – 13], известные методы, как правило, не являются универсальными, имеют ограничения применимости в различных режимах и сложных условиях выполнения наблюдений. Выполненный анализ основных известных подходов показывает, что методы, алгоритмы и программное обеспечение обнаружения, оценки (фиксации) и устранения фазовых скачков должны удовлетворять следующим общим требованиям:

- программное обеспечение должно выполнять обработку данных отдельных ГНСС станций (приемников), также как и выполнять аналогичную обработку традиционных одинарных и двойных разностей наблюдений "станция-станция", "спутник-спутник";
- входные данные – RINEX–файлы двухчастотных или одночастотных "сырых" ГНСС–наблюдений; программное обеспечение предварительной обработки должно выполнять обработку данных стандартного ряда темпов их регистрации 1/30 – 1 *Гц*;
- в ходе обработки должны быть выполнены обнаружение, оценка и коррекция скачков

наблюдений диапазонов частот ГНСС–сигналов L_1 и L_2 (или только L_1), которые были получены в статическом и в кинематическом "stop&go" режимах геодезической съемки, для каждой радиолинии (трассы) "спутник – приемник" отдельно.

С учетом указанных условий автором предложен новый универсальный подход, который базируется на результатах анализа и результатах многочисленных экспериментальных исследований и прошел успешную апробацию [13, 14]. Предложенный метод и алгоритмы решения задачи восстановления непрерывности фазовых наблюдений используют не только набор новых математических приемов обработки, но также и известные проверенные практикой процедуры. Для исключения фазовых циклических (и полуциклических) скачков одночастотных и двухчастотных ГНСС–наблюдений предлагаемый подход реализует интервальную итерационную процедуру обнаружения и оценки (фиксации) скачков одинарных разностей фаз между парами спутников с последующей их идентификацией на отдельных трассах "спутники – приемник". Последующая идентификация фазовых скачков на отдельных трассах "спутники – приемник" выполняется по результатам анализа скачков во всех возможных наборах линейных комбинаций фазовых наблюдений (одинарных разностях фаз между спутниками). После восстановления непрерывности L_1 наблюдений подобная обработка применяется и к линейным так называемым "безгеометрическим" комбинациям ($L_1 - L_2$) для обнаружения и устранения скачков в L_2-наблюдениях.

Этот метод, как показали исследования, особенно эффективен для обработки статических наблюдений, а его модификация позволяет решить задачу восстановления непрерывности и кинематических фазовых наблюдений. Для обработки разреженных наблюдений (темп данных – 1/5 – 1/30 *Гц*) перманентных референцных IGS/EPN ГНСС станций предложено использовать все возможные двойные разности наблюдений с целью исключить вариации часов спутников в увеличенных интервалах между эпохами и увеличить надежность обнаружения и точность оценки скачков фаз [12]. Адаптация предложенного метода и алгоритмов для совместной обработки фазовых ГНСС (EGPS/ГЛОНАСС/EGNOS/ Galileo/Compass) наблюдений ввиду их идентичности не представляет принципиальных затруднений. В частности, данный метод был успешно протестирован на реальных измерениях EGNOS и ГЛОНАСС. Алгоритмы устранения фазовых скачков реализованы в программном MATLAB-комплексе «OCTAVA_PPA» предварительной обработки и анализа ГНСС–наблюдений [14].

Тестирование предложенного метода и реализующих его алгоритмов позволило определить и ограничения их применимости. Это касается допустимых интервалов пропусков наблюдений, уровня шумов и т.д. и оптимальных параметров настройки процедур обработки для каждого типа наблюдений. Однако главное ограничение рассматриваемого метода заключается в снижении надежности обнаружения и фиксации скачков наблюдений, которые выполнялись в сложных условиях затенений и ограничений радиовидимости, особенно в условиях городских застроек, высотных зданий и др. Это приводит к частым разрывам слежения и появления кратковременных (до 5 – 30 *с*) пропусков и даже полной потери наблюдений. Из-за этого происходит дробление интервалов наблюдений на отдельные участки, что на последующих этапах (РФН и позиционирования) приводит к уменьшению вероятности отыскания правильного целочисленного фазового решения задачи РФН и, соответственно, к аномальным решениям задачи определения координат.

Расширить возможности и повысить надежность устранения фазовых скачков одночастотных и двухчастотных кинематических наблюдений в сложных условиях позволяет применение другого более эффективного и надёжного альтернативного метода восстановления непрерывности фазовых наблюдений (обнаружения, оценки и устранения дискретных фазовых скачков), разработанного для условий полной или частичной кратковременной потери наблюдений, позволяющего в значительной степени парировать воздействие дробления интервалов наблюдений на отдельные участки на качество решения задачи позиционирования. Описанный альтернативный метод устранения фазовых скачков (детально изложен в работах [15, 16]) основан на использовании приращений наблюдений по времени и специальной про-

цедуры совместного МНК–оценивания совокупности континуальных (приращений координат и расхождений часов) и дискретных параметров (циклических фазовых скачков). Оценивание циклических фазовых скачков осуществляется прямым перебором в предварительно определённой области поиска по аналогии с тем, как выполняется оценивание дискретных параметров в процедурах РФН. Тестирование предложенных методов и алгоритмов проводилась с использованием реальных статических и кинематических ГНСС наблюдений, полученных в рамках совместного инновационного проекта ГАО НАНУ и ХНУРЭ в 2007 – 2010 гг., а также GPS–наблюдений, выполненных на подвижных носителях (вертолет, самолет) при проведении аэрофотосъемок.

Цель данной статьи – описание и обобщение результатов создания и тестирования (верификации) новых методов и алгоритмов надежного решения задачи обнаружения, оценки и устранения фазовых скачков статических и кинематических ГНСС–наблюдений, а также условий их эффективного практического применения.

Концепция решения задачи обнаружения и устранения скачков фазовых наблюдений

Разработанная концепция предполагает использование наблюдений отдельной ГНСС–станции/приемника (темп данных 1 – 1/5 *Гц*) и выполняет решение задачи коррекции фазовых циклических скачков в несколько этапов.

На предварительном этапе после преобразования данных из формата RINEX в удобную для дальнейшей обработки структуру реализуются стандартные навигационные вычисления с использованием кодовых наблюдений. Затем выполняется интерполяция всех наблюдений (кодовых, фазовых, доплеровских) на целые секунды ГНСС путем вычисления производной псевдоскорости, если в составе RINEX–файле «сырых» наблюдений отсутствовали доплеровские наблюдения. На предварительном этапе также выполняется первое редактирование выборок наблюдений (по углу места и длительности пропусков наблюдений).

Первый этап обработки базируется на следующем эмпирическом принципе:

Наиболее удовлетворительной функцией для обнаружения и оценки (фиксации) фазового скачка в какой либо точке этой функции является функция «без динамики», либо функция, в которой с высокой точностью «динамика» скомпенсирована каким-либо образом, например, с использованием дополнительной информации. Если уровень аддитивных шумов и коррелированной помехи (многолучевость и др. источники) заметно ниже величины искомого скачка, то, взяв первые разности по времени (приращения) от такой функции, можно достаточно просто обнаружить, оценить и устранить все одиночные либо множественные (последовательные) скачки, если они идентифицированы по принадлежности к наблюдениям диапазонов навигационных сигналов L_1 или L_2. Важно в этом случае, что бреши (пропуски наблюдений) не влияют на качество определения (обнаружения и оценки) скачков.

Многочисленные экспериментальные исследования показали, что для целей определения и устранения фазовых циклических скачков вполне пригодны одинарные разности фазовых наблюдений «спутник – спутник» $[(L_1 - L_1)_{ij}, (L_2 - L_2)_{ij}, (L_{WL} - L_{WL})_{ij}, (L_1 - L_2)_i, (L_1 - L_2)_j]$, т.к. в этих линейных комбинациях устраняются погрешности часов приемника, а шумы бортовых атомных часов спутников не превышают ~(5 – 7) *см.* При формировании таких разностей важна точная временная привязка наблюдений (см. предварительный этап обработки) из-за достаточно высокой скорости изменения обрабатываемых параметров.

Главная задача заключается в корректном способе исключения «динамики» обрабатываемых функций в условиях пропусков наблюдений, во многих случаях – протяженных. Обработка начинается с определения участков видимости референцных спутников, совместно покрывающих весь интервал наблюдений. На каждом участке формируются и обрабатываются все указанные разностные ЛК наблюдений каждого из спутников относительно выбранного референцного. В результате интервальной итерационной обработки обнаруживаются и оцениваются скачки одинарных разностей «спутник-спутник» наблюдений L_1 для каждой пары спутников. Далее производится «поточечный» анализ скачков для всего участка

наблюдений. При этом используется принцип трех совпадений из m разностей ($m \geq 4$; $(m+1)$ – количество наблюдаемых спутников в анализируемый момент). Если в какой-либо момент времени в трех разностях фаз (ЛК) из m комбинаций скачки отсутствуют, то с высокой вероятностью в этот момент они отсутствуют в наблюдениях фазы на всех четырех трассах «спутники – ГНСС-приемник». Если же в каких-либо трех ЛК из m обнаружены и оценены скачки одной и той же величины, то с высокой вероятностью принимается решение о наличии скачка в этот момент в наблюдениях референцного спутника. В результате этого анализа формируются оценки скачков отдельно для каждой трассы («i-й спутник – приемник» и «j-й спутник – приемник»). Такой подход может быть успешно применен как к одинарным разностям («станция – станция»), так и к двойным разностям фазовых наблюдений.

После полного устранения L_1 фазовых скачков, в случае двухчастотных наблюдений подобная обработка осуществляется применительно к «безгеометрической» ЛК (L_1–L_2) отдельно для каждой трассы «спутник – приемник». В результате оцениваются и исключаются фазовые скачки наблюдений и диапазона L_2.

Важная отличительная особенность и новизна предлагаемого подхода состоят в способе обработки тех или других линейных комбинаций наблюдений, в частности, речь идет о прямых одночастотных одинарных разностей «спутник – спутник» и о двухчастотных «безгеометрических» ЛК (L_1–L_2). В ходе вычислений используется только обрабатываемая функция (ЛК) и ее приращения по времени (аналог доплеровских наблюдений). Эти приращения сглаживаются (с использованием кубических сплайнов) с одновременным устранением всех выбросов (эффекты скачков, пропусков наблюдений, аномальных измерений) и последующей интерполяцией сглаживающей функции на все пропущенные эпохи. Затем формируется разность между начальной функцией и проинтегрированной по времени «бесскачковой» сглаживающей приращения функцией. Полученная новая остаточная функция после одной-двух итераций сохраняет начальные скачки и становится близкой к ступенчатой (при наличии скачков) с исключенными эффектами влияния пропусков наблюдений. Применение одинарных приращений по времени к такой новой функции позволяет (при шумовом уровне, который не превышает выбранного порога – доли длины волны фазовых наблюдений) достаточно просто обнаружить и точно зафиксировать (например, оценить целое число длин волн в скачке) все скачки. На рис. 1 приведена иллюстрация разработанного способа оценки и устранения фазовых скачков в виде последовательности действий. Надежность обнаружения и оценки фазовых скачков зависит от того, насколько точно возможно выполнить аппроксимацию приращений по времени исходной функции. Это зависит от ряда факторов: соответствия динамики обрабатываемой функции темпу съема наблюдений, протяженности пропусков наблюдений, их шумового уровня, частоты возникновения пропусков и скачков.

Испытания с использованием статических наблюдений, выполненных в различных условиях, доказали работоспособность и эффективность предложенного метода по сравнению с известными зарубежными аналогами. Программное обеспечение, построенное на основе этого метода, позволяет выполнять успешное устранение фазовых циклических и полуциклических скачков, включая начальные и конечные участки выборок наблюдений с низкими углами места, и в условиях возникновения непрерывных скачков и пропусков наблюдений без использования дополнительной информации.

Описанный метод обработки имеет заметные преимущества перед аналогами [6 – 9]. Во-первых, формирование приращений начальной функции значительно уменьшает ее динамику и соответственно точность процесса аппроксимации приращений по сравнению с прямым сглаживанием начальной функции. Во-вторых, после завершения повторяющейся процедуры компенсации начальной динамики исходной функции для обнаружения и фиксации всех скачков используются только одинарные приращения. Дополнительные сложные и не всегда надежные процедуры для разделения скачков не требуются. В-третьих, технология универсальна, применима и к одночастотным и к двухчастотным статическим наблюдениям.

Повышение надежности предложенного метода может быть достигнуто за счет его совместного использования с другими методами и использования дополнительной информации

(доплеровские наблюдения, бортовые эфемериды спутников, «кооперативная» обработка пропущенных наблюдений для части спутников рабочего созвездия и т.д.). В обеспечение этого разработаны вспомогательные методы увеличения надежности, которые также протестированы с использованием наблюдений, полученных в режиме кинематических измерений. Адаптация предложенного метода и алгоритмов для совместной обработки фазовых наблюдений нескольких ГНСС (GPS/ГЛОНАСС/EGNOS/Galileo/Compass), как показали эксперименты с наблюдениями ГЛОНАСС и EGNOS, не представляет принципиальных трудностей.

Ниже представлены иллюстрации (рис. 2, 3) результатов определения и коррекции циклических фазовых скачков с использованием предложенных метода и алгоритмов применительно к наблюдениям приемника NovAtel OEM4-G2 [11, 14] для GPS спутников SV9 (β=5° – 10°) – рис. 2, SV26 (β=5° – 23°) – рис. 3. На рисунках все представленные значения функций даны в метрах, время – в секундах. Последовательность рисунков следующая. На рис. 2, *а – е*, рис. 3, *а – е* представлены двухчастотные (L_1 и L_2) ЛК фазовых и кодовых наблюдений, чувствительных к фазовым скачкам. Рис. 2, *а, б, в*, рис. 3 *а, б, в* соответственно показывают эти ЛК *перед* устранением фазовых скачков, рис. 2, *г, д, е*, рис. 3, *г, д, е – после* устранения фазовых скачков. На рис. 2, *а, б*, рис. 3, *а, б* (*до обработки*), рис. 2, *г, д*, рис. 3, *г, д* (*после обработки*) даются линейные комбинации ($L_1 - L_2$) (рис. 2, *а, г*, рис. 3, *а, г* и их приращения по времени – рис. 2, *б, д*, рис. 3, *б, д*).

Эти рисунки ясно показывают фазовые скачки (рис. 2, *а*, рис. 3, *а*), их появление в приращениях по времени (рис. 2, *б*, рис. 3, *б*) и шумы фазовых наблюдений (рис. 2, *д*, рис. 3, *д*). Рис. 2, *а, б, в*, рис. 3, *а, б, в* (до обработки), рис. 2, *г, д, е*, рис. 3, *г, д, е* (после обработки) иллюстрирует «безионосфесную» и «безгеометрическую» линейные комбинации, функцию четырех наблюдений F. Эта уникальная ЛК чувствительна к разности L_1 и L_2 фазовых циклических скачков с длиной волны λ_{WL}= 86,2 *см*. Линейная комбинация F является комплиментарной к разности (L_1–L_2) в случае возникновения двойных фазовых скачков (на одну и ту же эпоху), включая критические комбинации циклических скачков, когда $\Delta N_1 \lambda_1 - \Delta N_2 \lambda_2 \approx 0$. Здесь – ΔN_1, ΔN_2 – циклические скачки, а λ_1, λ_2 – длины волн несущих навигационных сигналов L_1 и L_2 соответственно [10, 11, 13].

Тестирование и представленные результаты демонстрируют успешное восстановление непрерывности статических фазовых наблюдений (с темпом 1 – 1/5 *Гц*) отдельной станции (приемника) с использованием разработанного подхода. Созданное программное обеспечение позволяет обнаруживать, определять и устранять фазовые циклические скачки даже на сложных участках (низкие углы места β=5° – 10°) и при повышенном уровне шумовых погрешностей, с многочисленными слипами и протяженными пропусками наблюдений.

Предложенный подход был также применен для реализации алгоритмов и программных модулей обработки разреженных статических наблюдений (типовой темп 1/30 *Гц*) и кинематических наблюдений (темп 1 *Гц* и выше). Эти алгоритмы подробно описаны ниже.

В ходе обработки разреженных наблюдений перманентных референцных IGS/EPN станций используются все возможные двойные разности наблюдений с целью исключить вариации часов спутников в увеличенных интервалах между эпохами и увеличить надежность обнаружения и точность оценки скачков фаз [12].

При обработке наблюдений с высокой динамикой перед использованием основного метода обнаружения, оценки и устранения фазовых скачков применяется предварительная процедура, позволяющая оценить и компенсировать динамику наблюдений на трассах «спутники – ровер». Для этой цели выполняется точная оценка приращений координат и вариаций часов приемника с использованием приращений фаз на соседних эпохах наблюдений с параллельной отбраковкой возможных фазовых скачков и выбросов.

Описанные подход и алгоритмы обработки реализованы в программном MATLAB-комплексе «OCTAVA_PPA» для обработки и анализа ГНСС-наблюдений [14].

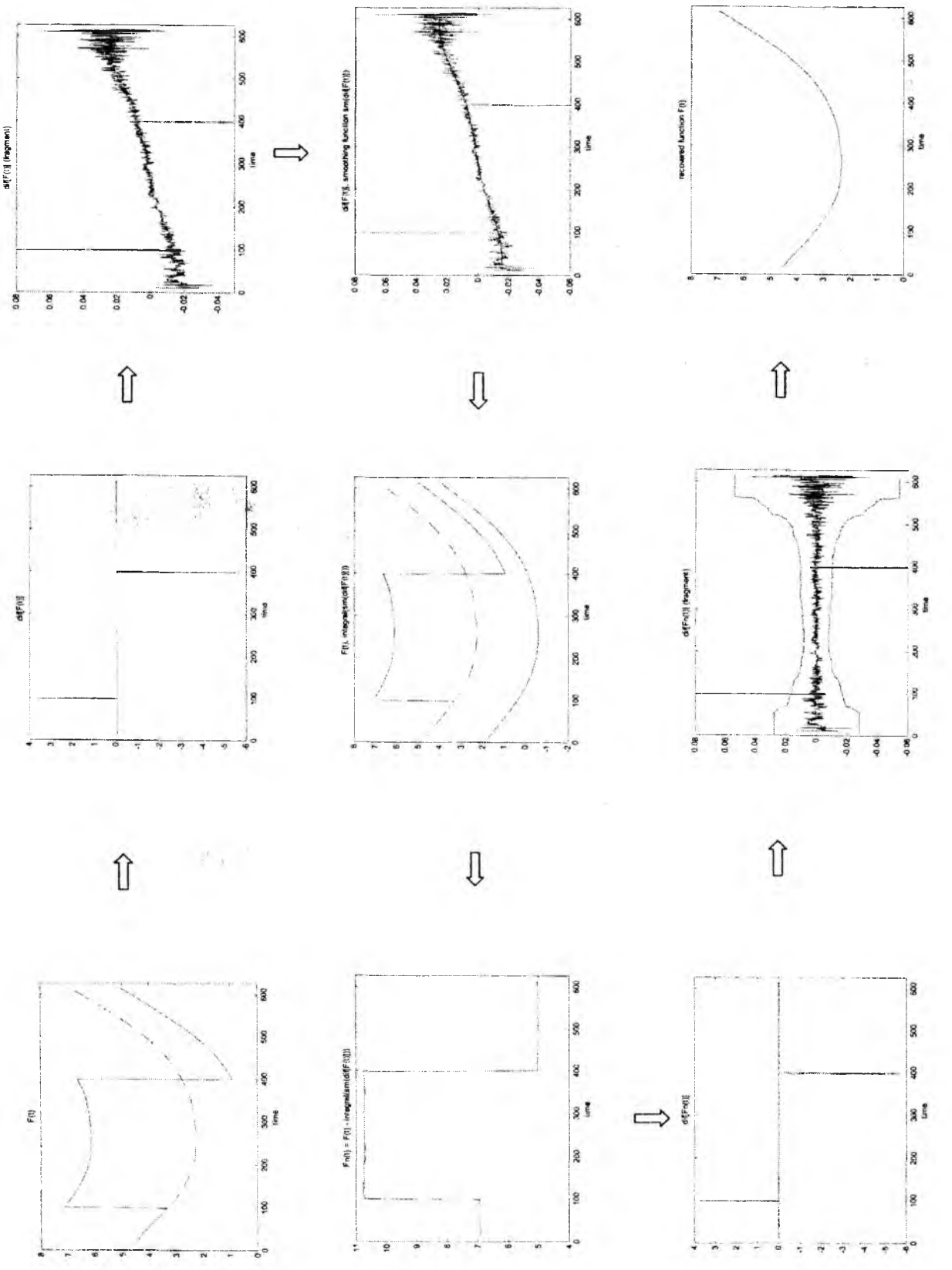
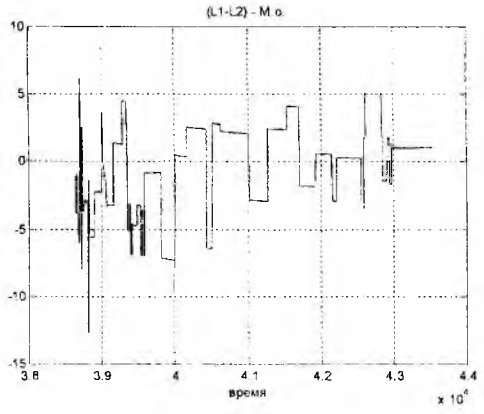
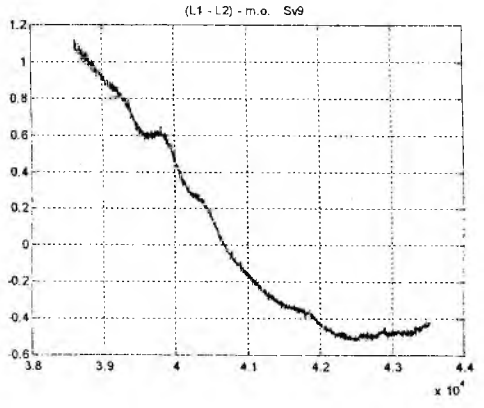


Рис. 1



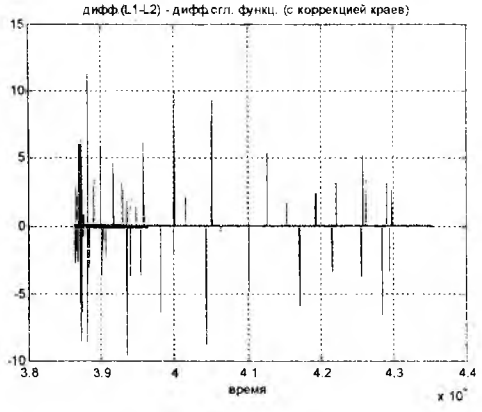
до обработки

а



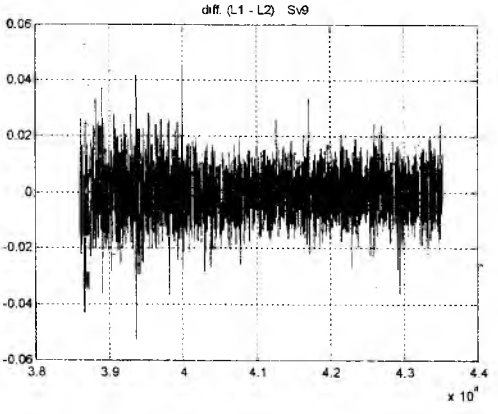
после обработки

г



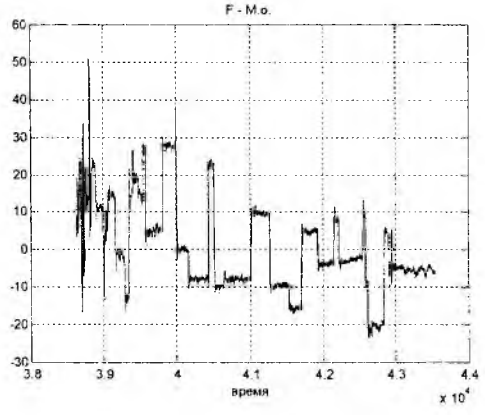
до обработки

б



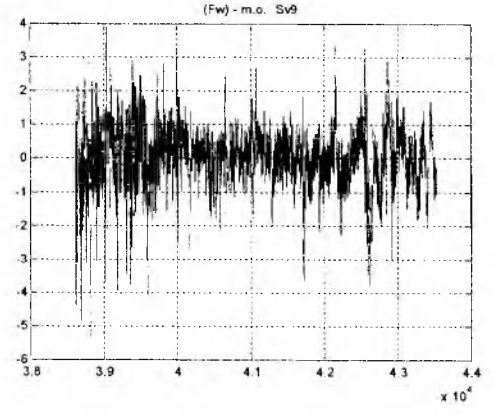
после обработки

д



до обработки

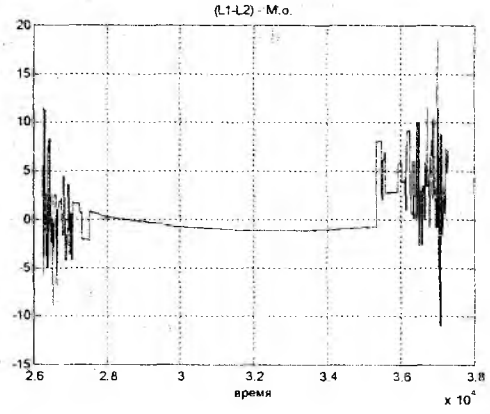
в



после обработки

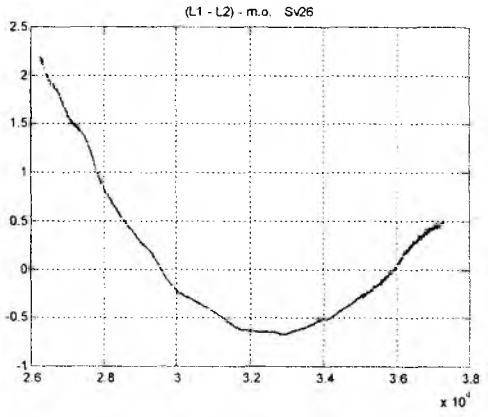
е

Рис. 2



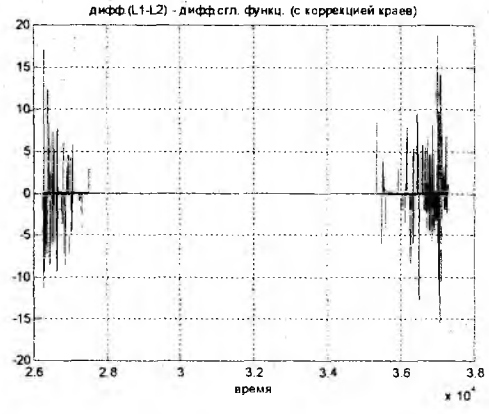
до обработки

а



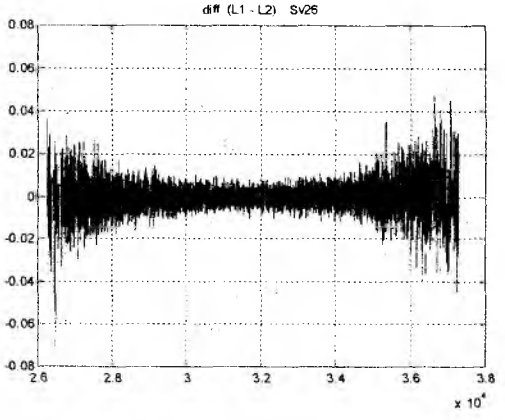
после обработки

г



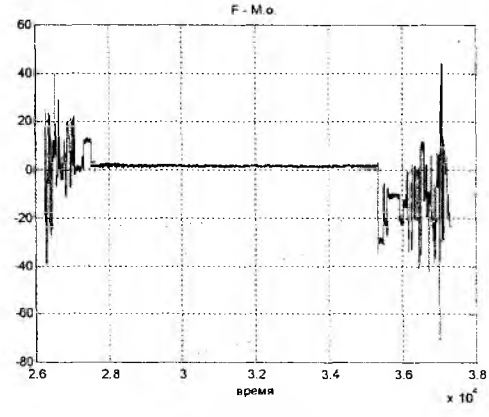
до обработки

б



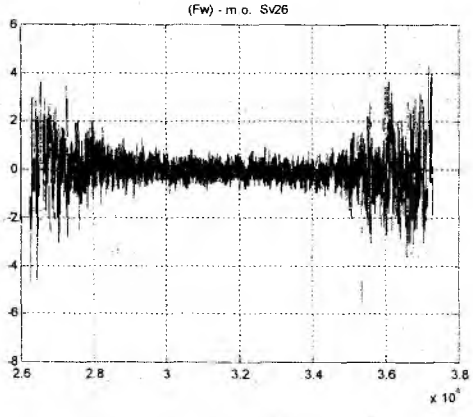
после обработки

д



до обработки

в



после обработки

е

Рис. 3

Особенности и ограничения точной коррекции фазовых скачков

Тестирование предложенного метода позволило определить ограничения его применимости. Это касается допустимых интервалов пропусков наблюдений, уровня шумов и оптимальных параметров настройки процедур обработки для каждого типа наблюдений.

Одночастотные одинарные разности фазовых наблюдений дают возможность обнаруживать и оценивать фазовые скачки наблюдений отдельной станции, однако применение

этой линейной комбинации имеет ограничение. Величина допустимого интервала пропуска наблюдений, при котором гарантируется надежное выявление скачка, для одинарных разностей зависит от различных факторов (ионосферных вариаций, шумовых характеристик часов спутников и др.). Как показали экспериментальные исследования, для темпа обрабатываемых данных 1 *Гц* допустимые интервалы пропусков наблюдений не должны превышать 50 – 70 отсчетов. В случае превышения допустимых интервалов необходимо отделить интервалы наблюдений спутника и выполнять обработку для каждого участка отдельно. Это главное ограничение по использованию предложенных алгоритмов обработки. Кроме того, проблемными для обработки являются участки (как правило, при наблюдениях на «низких» углах места) с чередующимися пропусками. Особенно это критично для краевых участков обрабатываемой выборки. Для повышения надежности обработки рекомендуется делать предыдущее редактирование наблюдений – выделять непрерывные (с допустимыми интервалами пропусков) участки, отсекать "плохие" краевые участки.

Наилучшим критерием тестирования качества обработки наблюдений в части исправления циклических фазовых скачков является использование двойных разностей наблюдений (между парами станций и парами спутников), так как эти линейные комбинации имеют меньшую динамику и не зависят от уходов часов приемников и спутников. Использование двойных разностей фазовых наблюдений позволяет реализовать надежную обработку даже сильно разреженных (1/30 *Гц*) наблюдений перманентных станций (см. ниже). Для реализации такой обработки необходимо использование наблюдений нескольких (сети) станций.

Однако главное ограничение рассматриваемого метода заключается в снижении надежности обнаружения и фиксации скачков наблюдений, которые выполнялись в сложных условиях затенений и ограничений радиовидимости, особенно в условиях городских застроек, высотных зданий и др. Это приводит к частым разрывам слежения и появлению кратковременных (до 5 – 30 *с*) пропусков и даже полной потери наблюдений. Из-за этого происходит дробление интервалов наблюдений на отдельные участки, на которых фазовые скачки можно устранить.

Расширить возможности и повысить надежность устранения фазовых скачков одночастотных и двухчастотных кинематических наблюдений в сложных условиях позволяет применение другого более эффективного и надёжного метода восстановления непрерывности фазовых наблюдений (обнаружения, оценки и устранения дискретных фазовых скачков) для условий полной или частичной кратковременной потери наблюдений. Кратко описанный ниже альтернативный метод устранения фазовых скачков [15, 16] основан на использовании приращений наблюдений по времени и специальной процедуры совместного МНК-оценивания совокупности континуальных (приращений координат и расхождений часов) и дискретных параметров (циклических фазовых скачков). Циклические фазовые скачки оцениваются прямым перебором в предварительно определённой области поиска по аналогии, так же, как оцениваются дискретные параметры в процедурах РФН.

Обнаружение, оценка, идентификация и устранение фазовых скачков разреженных наблюдений

Как показали исследования [12], обработку разреженных наблюдений (темп данных 1/30 – 1/10 *Гц* – для перманентных референцных IGS/EPN станций) с использованием разработанного метода выявления и исключение фазовых скачков необходимо выполнять с использованием двойных разностей фазовых наблюдений сети станций. В таблице представлены результаты определения предельных значений интервалов пропущенных наблюдений, при которых возникают ошибочные решения и снижается надежность устранения фазовых прыжков. Кратко описана модификация предложенных выше метода и алгоритмов обработки применительно к разреженным наблюдениям и процедура идентификации и исключения фазовых скачков на отдельных радиолиниях «спутники – станции».

В ходе выполненных исследований было выявлено, что компенсация в одинарных разностях фазовых наблюдений (между спутниками) расчетных расстояний (геометрическая составляющая), частотно-временных уходов спутниковых часов и тропосферных задержек дает возможность визуализировать и оценить уровень флуктуаций фаз опорных генераторов спутниковых часов. Было установлено, что эти флуктуации ограничивают допустимое число пропущенных эпох измерений. При темпе наблюдений 1/30 *Гц* пропуск всего одной точки наблюдений приводит к пропуску 60 *с* наблюдений. Если пропуск наблюдений находится на участке с большой крутизной, то интерполировать пропущенные точки с необходимой точностью сложно. Проявить скачок фазы в один-два цикла при таких условиях возможно только при совместной обработке одночастотных наблюдений нескольких перманентных станций. Учет динамики движения спутников, уходов спутниковых часов и тропосферной задержки не дают выигрыша в надежности выявления и оценки фазовых скачков. Основным фактором, который влияет на качество обработки одинарных разностей фазовых наблюдений, являются флуктуации бортовых часов спутников. В таблице приведены допустимые интервалы пропусков наблюдений для исследованных линейных комбинаций, обеспечивающих вероятность обнаружения и оценки фазового циклического скачка ~99% [12].

Линейная комбинация	Требует двухчастотных наблюдений	Используемые для обработки наблюдения	Интервал допустимых пропусков наблюдений на краю выборки, *с*	Интервал допустимых пропусков наблюдений в центре выборки, *с*
Одинарные одночастотные (L_1-L_1; L_2-L_2) разности	–	одиночная станция	без пропусков вероятность принятия правильного решения (обнаружения и оценки скачка) ~50%	
Одинарные (между спутниками) Wide-Lane разности	Да	одиночная станция	120 – 150	180 – 210
Одинарные ((L_1-L_1)-(L_2-L_2)) разности (между спутниками)	Да	одиночная станция	120 – 150	180 – 210
Двойные разности (между станциями и спутниками)	Нет	несколько (сеть) станций	120 – 150	210 – 270

В результате исследований получены следующие выводы:

- при обработке одинарных разностей фаз между навигационными спутниками одной перманентной станции и темпе информации 1/30 *Гц* невозможно гарантировать надежную оценку фазового циклического прыжка из-за погрешности интерполяции флуктуаций спутниковых часов даже при отсутствии пропусков наблюдений;
- на этапе подготовки данных к обработке линейных комбинаций наблюдений и при первичном редактировании данных необходимо определить участки, на которых продолжительность пропусков наблюдений превышает рекомендованные величины. Участки наблюдений, которые содержат пропуски, и эти пропуски превышают допустимые величины, должны обрабатываться отдельно;
- наиболее эффективна обработка разреженных наблюдений не одной станции, а их совместная (сетевая) обработка. При этом возможна надежная обработка как одночастотных, так и двухчастотных наблюдений с темпом данных 1/30 *Гц*. Для упрощения логики идентификации перманентной станции и радиолинии, где произошел фазовый скачок, рекомендуется совместная обработка наблюдений перманентных станций с использованием наблюдений

хотя бы одной станции с темпом данных 1 – 1/5 *Гц*. После устранения скачков фазовых наблюдений станции с темпом данных 1 *Гц* двойные разности L_1-наблюдений перманентных станций относительно этой станции позволяют легко обнаружить, оценить и идентифицировать скачки в наблюдениях перманентных станций;

- после восстановления непрерывности наблюдений на частоте L_1, линейная комбинация (L_1–L_2) дает возможность надежно обнаружить и устранить скачки в L_2-наблюдениях на каждой радиолинии отдельно. При реализации данных рекомендаций возможно увеличение допустимой продолжительности пропусков наблюдений, при которых обеспечивается надежная оценка фазовых скачков (вероятность правильного выявления 99 %) с 30 до 120 – 150 *с*;

- при обработке двухчастотных наблюдений отдельной перманентной станции с темпом информации 1/30 *Гц* восстановление непрерывности фазовых наблюдений возможно с использованием двух линейных комбинаций (например, Wide-Lane и (L_1–L_2) [1]). После обнаружения и идентификации фазовых циклических скачков в каждой из линейных комбинаций производится совместная обработка полученных данных для идентификации трасс («спутник – приемник») наблюдений, в которых произошли фазовые скачки.

Модифицированная процедура обработки фазовых разреженных наблюдений сети перманентных референцных станций сводится к следующим действиям.

Исходная информация:

1) фазовые наблюдения L_1 должны быть поделены на участки непрерывности с учетом заранее оговоренных условий длительности пропусков, количества точек (без пропусков) по краям, общей длительности участка, количества точек между пропусками, общего количества точек на участке;

2) заранее должно быть осуществлено ранжирование (определение приоритетности) станций, исходя из количества наблюдений L_1, при этом первая станция должна быть, по возможности, с наилучшими данными.

Рекомендуется вводить в состав сети станций в качестве первой – станцию, которая уже прошла полный пре-процессинг наблюдений с темпом 1 – 1/5 *Гц*.

Рекомендуемые критерии ранжирования станций по оценке качества наблюдений:

1) отсутствие пропусков эпох наблюдений – одновременно всех (C/A, P_2, L_1, L_2) наблюдений по всем спутникам (пропуски более одной эпохи); каждый такой пропуск образует новые сессии наблюдений;

2) отсутствие длительных (120 *с* и более) пропусков – участков наблюдений для каких-либо (одного или нескольких) спутников (не одновременно для всех) на углах места свыше 15^0;

3) минимальное число отбракованных наблюдений L_1, L_2;

4) минимальные потери L_1 наблюдений после деления на участки по сравнению с расчетной длительностью (по заданным ограничениям по углу места или по интервалу измерений для каждого спутника).

Обработка кинематических фазовых наблюдений

При обработке наблюдений, полученных в условиях высокой динамики (кинематическая съемка), когда темп наблюдений (обычно 1 *Гц*) не соответствует спектру процесса движения объекта, применять только описанные выше интерполяционные схемы невозможно. Это относится к геодезическим наблюдениям в движении.

Суть алгоритма предварительной обработки наблюдений с высокой динамикой заключается в точной оценке текущих изменений (приращений) координат движущегося объекта и формировании разностей измеренных фаз принятых навигационных сигналов и расчетных их значений, в которых динамика движения была бы практически исключена. Это позволяет воспользоваться ранее описанным интерполяционным методом обработки фазовых наблюдений для выявления, оценки и исключения фазовых скачков статических наблюдений.

Как показали исследования, точность оценки приращений координат движущегося объекта и уходов часов приемника достигает уровня 0,5 – 2 *см*. Это позволяет рассчитать и практически целиком устранить динамику изменения фазовой задержки (геометрия, уходы часов, тропосферная и ионосферная задержки) на отдельных радиолиниях «спутник – приемник», визуализировать и анализировать кодовые и фазовые наблюдения по уровню многолучевости и шумов. Задача состоит в получении надежных оценок приращений координат и уходов часов роверного приемника. Приращение координат и времени можно проинтегрировать и с точностью до констант восстановить фазовые наблюдения и их ЛК, чтобы потом это использовать для обработки «квази-бездинамичных» фазовых наблюдений с целью выявления, оценки и починки скачков. Вторая задача состоит в выявлении моментов скачков в наблюдениях отдельных спутников из текущего рабочего созвездия в ходе решения навигационной задачи (по невязкам измеренных и расчетных значений наблюдений).

Таким образом, идея метода заключается в использовании наблюдений всех спутников в сопредельные моменты времени (где это возможно) сначала для оценки приращений координат объекта и затем – для устранения динамики объекта и спутников и последующей обработки полученных разностей фазовых наблюдений и предвычисленных значений для выявления и оценки (фиксации) скачков. Данная концепция объединяет в себе два подхода – оценку и компенсацию динамики движения объекта в наблюдениях и уже используемую интерполяционную схему обработки. Она может быть реализована также в итерационной схеме, когда после первой починки слипов повторная оценка приращений координат «ровера» будет более надежной.

Ниже представлено описание решения поставленной задачи.

Положим, что «роверный» (мобильный) движущийся приемник (R – «rover») и базовая станция (BS – «base station») в общем случае производят двух- или одночастотные кодовые и фазовые наблюдения. Фазовые наблюдения неоднозначны и подвержены скачкам. Задача состоит в том, чтобы получить надежные оценки приращений координат и уходов часов «ровера» (или «ровера» относительно базовой станции BS при реализации дифференциального режима измерений). Полагаем, что предварительная обработка наблюдений BS уже выполнена, и в фазовых наблюдениях BS нет скачков. Приращения координат и времени можно проинтегрировать и с точностью до констант восстановить фазовые псевдодальности и их комбинации, чтобы затем это использовать для обработки «квази-бездинамичных» фазовых наблюдений с целью обнаружения, оценки и починки фазовых скачков. Параллельно необходимо решить задачу предварительного обнаружения моментов фазовых скачков в наблюдениях отдельных спутников из текущего рабочего созвездия в ходе решения навигационной задачи (по невязкам измеренных и расчетных наблюдений).

Таким образом, идея метода состоит в использовании наблюдений всех спутников в смежные моменты времени (где это возможно) вначале для оценки приращений координат объекта и затем – для устранения динамики объекта и спутников и последующей обработки полученных разностей фазовых наблюдений и предвычисленных значений для обнаружения и оценки фазовых скачков. В ходе решения можно использовать предварительно обработанные и кодовые наблюдения (их приращения) совместно с приращениями фаз с учетом весов этих сильно неравноточных наблюдений. Это позволит регуляризировать решение навигационной задачи (НЗ) и повысить надежность обнаружения выбросов приращений фазы, обусловленных скачками.

Полагаем, что как в наблюдениях «ровера», так и в наблюдениях базовой станции выделены участки наблюдений, на которых предполагается исключить фазовые скачки, кодовые псевдодальности непрерывны и вычищены от аномальных наблюдений, скачков часов, все наблюдения приведены к единой равнодискретной шкале времени ГНСС.

Рассмотрим уравнения наблюдений и их приращений (см. ниже), в которых опустим медленные аппаратурные изменения (задержки в измерительных трактах), изменения из-за неточного знания фазовых характеристик (фазовых центров и вариаций фазовых диаграмм)

передающих и приемных антенн, wind-up-эффекты, релятивистские и геодинамические эффекты. Полагаем также, что в приращениях наблюдений практически уходят погрешности координатной привязки базовой станции.

Полагаем, что при обработке наблюдений движущегося «роверного» приемника могут использоваться дифференциальные коррекции (фазовые наблюдения) базовой (ых) станции (й), при этом фазовые наблюдения базовой станции не должны содержать фазовых скачков.

Здесь также полагаем, что из-за большего количества циклических скачков в наблюдениях на частоте L_2 (по сравнению с наблюдениями на частоте L_1) на данном этапе обработки будем использовать только наблюдения диапазона L_1. Влияние медленно меняющихся погрешностей (эфемериды, ионосфера, тропосфера, медленные уходы часов спутников) на точность оценки приращений координат будет компенсироваться с использованием либо соответствующих моделей (допустимых для недифференциального режима интервалов формирования приращений наблюдений) либо с использованием дифференциального режима и тех же моделей. Обработка наблюдений диапазона L_2 (в части устранения фазовых скачков) должна проводиться по алгоритмам для статического режима наблюдений на основе использования «безгеометрических» разностей (L_1–L_2), как это было описано ранее.

Определение допустимых интервалов должно быть произведено путем расчетов с использованием моделей изменения эфемеридных погрешностей, тропосферной и ионосферной составляющих. Допустимый интервал должен быть таким, чтобы медленно меняющаяся суммарная компонента погрешности приращения фазы на этом интервале не превышала 2-3 *см*. Тогда будет возможно обнаружение, верная оценка и устранение циклических и даже полуциклических скачков фазы.

При последующем анализе уравнений наблюдений и источников и составляющих погрешностей будем анализировать интервалы оценки приращений координат от 1 до 100 *с*.

Полагаем, что при формировании последовательных приращений фазовых наблюдений для смежных моментов t_{k+1} и t_k имеем на эти моменты созвездие из m спутников ($m > 4$) и для момента t_k имеем хотя бы грубые (абсолютные) значения координат и расхождений шкал времени роверного приемника (для базовой станции координаты известны с высокой точностью).

Приращения кодовых и фазовых наблюдений на частоте L_1 (с учетом изложенного) для роверного («r») приемника или базовой станции («bs»):

$$\begin{cases} \Delta\hat{S}_c^j(t_{k+1},t_k) = \Delta\dot{R}^j(t_{k+1},t_k) + \Delta\Delta_{эф}^j(t_{k+1},t_k) + \Delta\Delta_{эф}^{GPS}(t_{k+1},t_k) - \\ -\Delta\Delta^{j,GPS}(t_{k+1},t_k) + \Delta T^j(t_{k+1},t_k) + \Delta\dot{I}^j(t_{k+1},t_k) + \delta\Delta\hat{S}_{mf}^j(t_{k+1},t_k); \\ \Delta\hat{L}^j(t_{k+1},t_k) = \Delta\dot{R}^j(t_{k+1},t_k) + \Delta\Delta_{эф}^j(t_{k+1},t_k) + \Delta\Delta^{GPS}(t_{k+1},t_k) - \\ -\Delta\Delta^{j,GPS}(t_{k+1},t_k) + \Delta T^j(t_{k+1},t_k) - \Delta\dot{I}^j(t_{k+1},t_k) + \xi_j \cdot \Delta N^j(t_{k+1},t_k)\cdot\lambda^* + \\ +\delta\Delta L_{mf}^j(t_{k+1},t_k), \end{cases} \quad (1)$$

где $\Delta\dot{R}^j(t_{k+1},t_k) = \dot{R}^j(t_{k+1}) - \dot{R}^j(t_k) = \sqrt{\underbrace{(x-\hat{X}^j)^2+(y-\hat{Y}^j)^2+(z-\hat{Z}^j)^2}_{t=t_{k+1}}} -$

$$- \sqrt{\underbrace{(x-\hat{X}^j)^2+(y-\hat{Y}^j)^2+(z-\hat{Z}^j)^2}_{t=t_k}};$$

x, y, z – неизвестные координаты «ровера» (x_r, y_r, z_r) на моменты t_{k+1}, t_k или известные координаты базовой станции $(\hat{x}_{BS}, \hat{y}_{BS}, \hat{z}_{BS})$.

В системе уравнений (1) опущены составляющие погрешностей приращений дальностей («BS – j-й SV») из-за остаточных погрешностей привязки базовой станции они пренебрежимо малы; в то же время в уравнениях для базовых станций (BS) целесообразно использовать «шляпки» над x_{BS}, y_{BS}, z_{BS}, не забывая, что это не точные значения координат, а их оценки;

$\Delta\Delta^j_{эф}$, $\Delta\Delta^{GPS}$, $\Delta\Delta^{jGPS}$, ΔT^j, ΔI^j – приращения эфемеридных погрешностей, приращения уходов часов приемника (величина, подлежащая оценке), приращения уходов часов спутников (погрешность) приращения тропосферной и ионосферной погрешностей соответственно;

$\xi_j \cdot \Delta N^j(t_{k+1}, t_k) \cdot \lambda^*$ – величина циклического ($\lambda^* = \lambda_1$) или полуциклического ($\lambda^* = \lambda_{1/2}$) скачка фазы L_1 при его возникновении ($\xi_j = 1$), ($\xi_j = 0$ – скачок отсутствует);

$\delta\Delta S^j_{mf}$, $\delta\Delta L^j_{mf}$ – погрешности из-за шумов и многолучевости.

При совместной обработке фазовых и кодовых приращений можно принять соотношение весов: $P_{\Delta L} = 100$, $(\sigma L \approx 1\ см)$; $P_S = 0{,}5 - 1$; $(\sigma S \approx 1 - 2\ м)$, размерность весов – $(1/м)$. При последующей обработке весовые соотношения можно уточнить, но на данном этапе для решений данной конкретной задачи это не столь важно.

Теперь видоизменим и проанализируем выражение для $\Delta\dot{R}^j(t_{k+1}, t_k)$. Введем обозначения искомых величин приращений координат, опустив для простоты обозначение момента времени t:

$$\Delta x = \Delta x_{k+1} = x_{k+1} - x_k;$$
$$\Delta y = \Delta y_{k+1} = y_{k+1} - y_k;$$
$$\Delta z = \Delta z_{k+1} = z_{k+1} - z_k,$$

где индекс k – означает момент t_k, а индекс $(k+1)$ – момент t_{k+1}.

Тогда

$$\Delta\dot{R}^j_{k+1} = \Delta\dot{R}^j(t_{k+1}, t_k) =$$
$$= \sqrt{\left[(\Delta x_{k+1} - \Delta\hat{X}^j_{k+1}) + (\hat{x}_k - \hat{X}^j_k)\right]^2 + \left[(\Delta y_{k+1} - \Delta\hat{Y}^j_{k+1}) + (\hat{y}_k - \hat{Y}^j_k)\right]^2 +}$$
$$+ \overline{\left[(\Delta z_{k+1} - \Delta\hat{Z}^j_{k+1}) + (\hat{z}_k - \hat{Z}^j_k)\right]^2} + \frac{\partial\Delta\dot{R}^j}{\partial(x_k, y_k, z_k)} \cdot \delta\begin{pmatrix} x_k \\ y_k \\ z_k \end{pmatrix}.$$

Здесь введены приближенные оценки координат «ровер»/базовая станция.

Упростим еще раз последнее выражение:

$$\Delta\dot{R}^j(t_{k+1}, t_k) = \Delta\dot{R}^j =$$
$$= \sqrt{\left[(\Delta x + (\hat{x}_R - \Delta\hat{X}^j_{R+1})\right]^2 + \left[(\Delta y + (\hat{y}_R - \hat{Y}^j_{R+1})\right]^2 + \left[(\Delta z + (\hat{z}_R - \hat{Z}^j_{R+1})\right]^2} -$$
$$- \sqrt{\left[\hat{x}_R - \Delta\hat{X}^j_{R+1}\right]^2 + \left[\hat{y}_R - \hat{Y}^j_{R+1}\right]^2 + \left[\hat{z}_R - \hat{Z}^j_{R+1}\right]^2} + \delta\Delta\dot{R}^j(\delta\vec{x}_k). \qquad (2)$$

Прежде чем приступить к анализу вариантов решения поставленной задачи оценки приращений координат и расхождения времени «роверного» приемника (или базовой станции) важно оценить возможность решения задачи с необходимым качеством, т.е. оценить погрешности и допустимые интервалы формирования и обработки приращений фазовых псевдодальностей.

Подразумевается, что требуемая точность оценки искомых приращений не должна быть хуже 9 *см* с учетом геометрического фактора (ГФ), который ограничивается типовыми значениями 2 – 4. Таким образом, суммарная погрешность определения приращений фаз не должна превышать 2 – 5 *см*. Это важно еще и для решения задачи надежного обнаружения «слипов» (до 0,5 λ) в приращениях фазы (фазовых псевдодальностей).

Очевидно, что точность приращений фазы при фиксированных флуктуационной и многолучевой составляющих погрешности будет зависеть от изменения медленно меняющихся составляющих:

эфемеридной;

ухода часов спутников;

абсолютных определений координат $\delta\vec{x}(t_k)$ в опорный момент t_k;

ионосферной;

тропосферной.

Рассмотрим эфемеридную составляющую погрешности и составляющую погрешности, обусловленную неточным определением абсолютных координат в начальный момент t_k.

Исходя из общих правил взятия дифференциала от выражения (2) можно получить соотношения для оценки рассматриваемых составляющих погрешностей приращений дальностей.

Нетрудно показать, что эфемеридная составляющая и составляющая из-за погрешности координат в момент t_k имеют идентичные (с точностью до знака) выражения для оценки погрешности приращений, т.е.:

$$\begin{cases} \delta\Delta\dot{R}^j(\delta\vec{x}^j_{эф}) \approx \left[\hat{\vec{a}}^j(t_{k+1}) - \hat{\vec{a}}^j(t_k)\right]^T \cdot \delta\vec{x}^j_{эф}; \\ \delta\Delta\dot{R}^j(\delta\vec{x}(t_k) \approx -\left[\hat{\vec{a}}^j(t_{k+1}) - \hat{\vec{a}}^j(t_k)\right]^T \cdot \delta\vec{x}(t_k). \end{cases}$$

Здесь $\delta\vec{x}^j_{эф}$ – вектор погрешностей эфемерид.

Предполагаем, что в орбитальной (связанной с местоположением спутника) системе координат *j*-го SV за интервал времени $\Delta t_{k+1} = (t_{k+1} - t_k)$ вектор $\delta\vec{x}^j_{эф}$ практически не изменяется (медленно меняющаяся погрешность) с учетом неизменности преобразования из орбитальной системы координат (СК) в ГСК (Гринвичская система координат) для почти круговой орбиты, но изменяется положение спутника и, соответственно, проекция $\delta\vec{x}^j_{эф}$ на направление "приемник –*j*-й SV" будет изменяться с учетом местоположения приемника на поверхности Земли;

$\delta\vec{x}(t_k)$ – вектор погрешностей координат приемника в момент t_k;

$\hat{\vec{a}}^j(t_k)$, $\hat{\vec{a}}^j(t_{k+1})$ – векторы направляющих косинусов в ГСК на трассе « *j* -й SV – приемник».

Рассмотрим плоскую (двумерную) модель (рис. 4) для оценки максимальной эфемеридной погрешности с учетом, что орбиты спутников GPS почти круговые, радиус Земли $R_З \approx 6400\ км$, высота орбит $H \approx 20000\ км$.

Легко показать, что $\delta\dot{R}^j_{эф} \approx \Delta_{эф} \cdot [\gamma_{k+1} - \gamma_k] = \Delta_{эф} \cdot \Delta\gamma(\phi)$, где $\Delta\gamma$ выражено в радианах. Это следует из выражения

$$\delta\dot{R}^j_{эф} = \partial\dot{R}^j_{эф}(t_{k+1}) - \delta\dot{R}^j_{эф}(t_k) = \Delta_{эф} \cdot cos(\alpha + \gamma_{k+1}) - \Delta_{эф}\, cos(\alpha + \gamma_k),$$

при условии достижения максимума ошибки: $\sin(\alpha + \gamma_k) + \dfrac{\Delta\gamma(\varphi)}{2} = 1$.

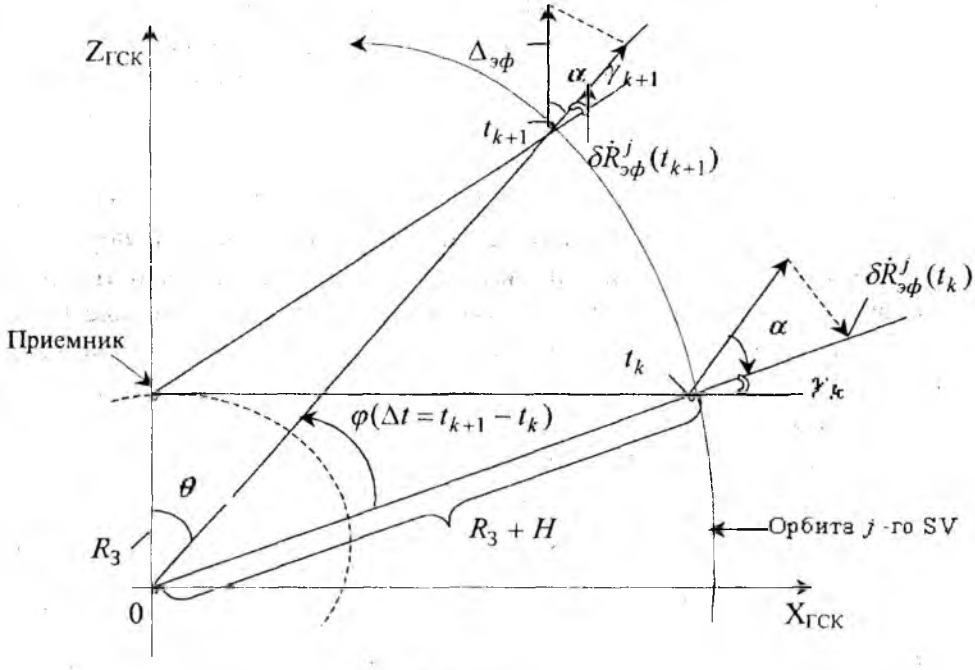


Рис. 4

Из двух треугольников на рисунке с использованием решений треугольников получим

$$tg\gamma_k = \frac{R_З \cdot sin(\theta + \phi)}{(R_З + H) - R_З \cdot cos(\theta + \phi)},$$

$$tg\gamma_{k+1} = \frac{R_З \cdot sin(\theta)}{(R_З + H) - R_З \cdot cos\,\theta},$$

откуда после взятия разности $tg\gamma_{k+1} - tg\gamma_k \approx \Delta\gamma(\phi)$ и нескольких преобразований имеем

$$\Delta\phi(\gamma) \approx \frac{cos\,\theta}{\left(1 + \dfrac{H}{R_З}\right) - cos\,\theta} \cdot \phi \approx \frac{cos\,\theta}{4 - cos\,\theta} \cdot \phi(\Delta t) \approx \frac{1}{3}\phi(\Delta t),$$

где φ – в радианах ($\varphi = 1{,}5 \cdot 10^{-4}(\Delta t = 1\ c) - 1{,}5 \cdot 10^{-2}\ рад(\Delta t = 100\ с)$),
тогда $\Delta\gamma^{(рад)}_{max} = 5 \cdot 10^{-5}(\Delta t = 1\ c) - 5 \cdot 10^{-3}(\Delta t = 100\ c)$ и
$\Delta\dot{R}^j_{эф_{max}} \approx 0{,}5\ мм \cdot (\Delta t = 1\ c) - 50\ мм\ (\Delta t = 100\ c)$ при $\Delta_{эф} = 10\ м$.

Аналогичные выводы можно сделать относительно составляющей из-за погрешностей координат приемника в момент t_k (при $\left|\delta \bar{X}\left(t_k\right)\right| \approx 10\,м$).

Таким образом, при обработке наблюдений фазы автономного приемника допустимый интервал не должен превышать 30 – 40 *с* с точки зрения вклада рассмотренных составляющих погрешностей.

Если речь идет о базовой станции, то с учетом ее точной координатной привязки погрешность $\delta \dot{R}^{j}_{\delta x} < 1\,мм$ при $\Delta t \le 100\,с$. Если применить дифференциальный метод и обрабатывать дифференциальную разность фаз, то даже на удалении роверного приемника от базовой станции до 1000 *км* погрешность эфемерид уменьшается в 20 – 30 раз, а погрешность определения координат ровера по кодовым наблюдениям может быть доведена до уровня 1 – 3 *м*, т. е. допустимый интервал разности фаз может быть увеличен до 100 – 300 *с*.

Теперь необходимо рассмотреть и оценить остальные медленно меняющиеся составляющие погрешностей разностей фазовых наблюдений – ухода часов спутников, ионосферной и тропосферной.

Что касается изменения уходов (медленных флуктуаций) $\Delta\Delta^{j,GPS}\left(t_{k+1}, t_k\right)$ часов спутников, то при обработке автономных наблюдений, как показывают экспериментальные исследования, допустимый интервал приращений фазы не должен превышать $\Delta t = 10 - 15$ *с*. При реализации дифференциального режима измерений (при этом фазовые измерения базовой станции должны быть очищены от скачков) рассматриваемая составляющая погрешностей наблюдений практически полностью компенсируется и допустимая длительность Δt будет зависеть от других факторов.

Для тропосферных и ионосферных составляющих необходимо провести отдельное исследование, особенно это требует тропосферная составляющая.

Предварительный анализ фазовых разностей (L_1–L_2), изменение которых во времени пропорционально $\left(k = \left(\gamma^2 - 1\right), \gamma = \frac{77}{60} \approx 1,28\right)$ изменению ионосферной задержки на частоте L_1, показал, что максимальные значения приращений ионосферной задержки на малых углах места 5 – 10 *угл. град.* составляют 0,1 – 02 *см/с*, а в некоторых случаях (при $\beta \approx 5^\circ - 7^\circ$) достигают значений 0,4 *см/с*. Это дает основание говорить, что при использовании и обработке автономных (одна станция/приемник) приращений наблюдений фазы (одной частоты) и при компенсации ионосферной составляющей с помощью модели (например, модели Клобушара) допустимый интервал Δt формирования и обработки приращений фазы L_1 не должен превышать $\Delta t_{k+1} \le 10 - 15$ *с*. Применение дифференциального режима наблюдений (вместе с модельной компенсацией) позволит заметно увеличить эффективность парирования ионосферной составляющей. Но из-за изменчивой и неоднородной природы ионосферы нужно признать, что при дифференциальном методе базовые линии не должны превышать 100 – 300 *км*.

При наличии двухчастотных наблюдений L_1 и L_2 можно организовать полную компенсацию ионосферной составляющей путем

а) формирования разностей (L_1–L_2);

б) получения гладкой равнодискретной функции приращений функции (L_1–L_2) в условиях наличия фазовых скачков L_1 и L_2 по известной и отработанной методике устранения динамики;

в) расчета (оценки) приращений ионосферной составляющей погрешности на требуемых интервалах Δt.

Такой двухчастотный подход может позволить практически полностью компенсировать ионосферную составляющую приращений фазовых наблюдений.

Что касается тропосферной составляющей, то очевидно, в приращения фазы обязательно нужно вводить модельные поправки (приращения поправок). Это позволит уменьшить тропосферную составляющую не менее, чем на порядок.

Приращение тропосферной погрешности, при условии, что «роверный» приемник или неподвижен или движется с относительно невысокой скоростью $v \le 60-100$ *км/ч*, можно выразить простой приближенной формулой

$$\delta\Delta\dot{R}^j_{\text{тр}} \approx 0{,}1 \cdot \frac{\partial T^j(\beta)}{\partial \beta} \cdot \Delta\beta ,$$

где $T^j(\beta)$ – тропосферная составляющая задержки сигнала; β – угол места j-го спутника относительно приемника; $\Delta\beta$ – изменение β за интервал Δt взятия приращений.

Величина $\Delta\beta$ для «ровера» не превышает 0,012 *угл. град./с*. Производная $\frac{\partial T(\beta)}{\partial \beta}$ для $\beta = 5^\circ$ не превышает величины 5 *м/угл. град*. Тогда максимальная величина тропосферной погрешности приращения фаз за $\Delta t = 1$ *с* составит 5 – 6 *мм* (при использовании модельной компенсации). Для углов $\beta \approx 10^\circ - 15^\circ$ $\delta\Delta\dot{R}^j_{\text{тр}} \approx 2-3$ *мм* (за секундный интервал). Эффективное уменьшение тропосферной погрешности с использованием дифференциального режима также ограничено базовыми расстояниями от 50 до 100 *км* в зависимости от различия условий измерений для «ровера» и базовой станции.

Таким образом, допустимый интервал Δt взятия приращений фаз в оговоренных условиях не должен превышать 5 – 10 *с* для автономных наблюдений и 15 – 30 *с* для дифференциальных. Для кинематических наблюдений характерен темп данных 1 *Гц* и выше. Поэтому рекомендуемый интервал приращений составляет 1 – 5 *с*, а при обработке наблюдений перманентных IGS-станций рекомендуется обрабатывать наблюдения приращений на интервале 10 – 30 *с* с реализацией дифференциального режима (сеть станций) с увеличением минимального угла места $\beta_{\min} \ge 10^\circ$ для уменьшения влияния атмосферных (тропосферных и ионосферных) погрешностей.

Теперь рассмотрим алгоритмы обработки одночастотных (L_1) фазовых наблюдений с целью оценить и устранить фазовые скачки в кинематическом режиме измерений. Обнаружение, оценка и исключение фазовых скачков фазовых наблюдений диапазона L_2 выполняются на основе обработки «безгеометрических» разностей ($L_1 - L_2$) с использованием алгоритмов, описанных выше для обработки статических наблюдений.

Система уравнений, связывающая приращения фазовых и кодовых наблюдений и искомых приращений координат фазового центра приемной антенны роверного приемника $\Delta x, \Delta y, \Delta z$ (в Гринвичской геоцентрической системе координат) и приращений Δ^{GPS} расхождения шкалы времени «ровера» относительно шкалы ГНСС (на примере GPS) (см. уравнения (1), (2)) на момент Δt_{k+1} относительно момента t_k может быть представлена в виде

$$\begin{cases} \Delta\hat{L}^j = \underbrace{\Delta\dot{R}^j + \Delta\Delta^{GPS}}_{F^j(\Delta t_k)} + \Delta N^j \cdot \lambda^* + \sqrt{\left[-\Delta\Delta^{j,GPS} + \Delta T^j - \Delta I^j\right]} + \delta\Delta L^j_{mf}; \\ \Delta\hat{S}^j = F^j\left(t_{k+1}\right) + \left[-\Delta\Delta^{j,GPS} + \Delta T^j - \Delta I^j\right] + \delta\Delta S^j_{mf}, \end{cases} \qquad (3)$$

где обозначения соответствуют выражениям (1) и (2), при этом индексы смежных моментов времени опущены;

$$\Delta\dot{R}^j = \sqrt{\left[(\Delta x + (\hat{x}_k - \Delta\hat{X}^j_{k+1})\right]^2 + \left[(\Delta y + (\hat{y}_k - \hat{Y}^j_{k+1})\right]^2 + \left[(\Delta z + (\hat{z}_k - \hat{Z}^j_{k+1})\right]^2} -$$
$$-\sqrt{(\hat{x}_k - \Delta\hat{X}^j_k)^2 + (\hat{y}_k - \hat{Y}^j_k)^2 + (\hat{z}_k - \hat{Z}^j_k)^2}.$$

Полагаем, что известны *на все моменты* времени обрабатываемого сеанса наблюдений:

- приближенные координаты x_k, y_k, z_k ровера и соответствующие оценки эфемерид всех спутников на моменты излучения (с учетом релятивистских коррекций и коррекций на вращение Земли); если пункт статический или это базовая станция, то $\vec{x}_k$ – известные постоянные значения;

- модели тропосферной и ионосферной коррекции, а также бортовые гладкие коррекции часов спутников;

- априорные средние значения среднеквадратической погрешности (СКП) шумовой и многолучевой составляющей погрешностей приращений наблюдений: $\sigma_{\Delta L} = 1{,}0\ см,\ \sigma_{\Delta S} = 100\ см$.

После расчета и ввода коррекций в наблюдения (3):

$$\begin{cases} \Delta\hat{\dot{L}}^j = \underbrace{\Delta\hat{L}^j + \Delta\hat{\Delta}^{j,GPS}}_{\hat{I}^j(\Delta t_k)} - \Delta\hat{T}^j + \Delta\hat{\dot{I}}^j \\ \Delta\hat{\dot{S}}^j = \Delta S^j + \Delta\hat{\Delta}^{j,GPS} - \Delta\hat{T}^j - \Delta\hat{\dot{I}}^j, \end{cases},$$

систему уравнений можно переписать как

$$\begin{cases} \Delta\hat{\dot{L}}^j = F^j(t_{k+1}) + \Delta N^j \cdot \lambda^* + \delta\Delta\hat{\dot{L}}^j_\Sigma; \\ \Delta\hat{\dot{S}}^j = F^j(t_{k+1}) + \delta\Delta\dot{S}^j_\Sigma, \end{cases}$$

где $\delta\Delta\dot{L}^j_\Sigma$, $\delta\Delta\dot{S}^j_\Sigma$ – *суммарные* остаточные после ввода коррекций погрешности и погрешности из-за многолучевости и шумов.

Или более детально, уравнения и их вес:

$$\left\{\begin{array}{l|l} p_{\Delta L} & \Delta\hat{\dot{L}}^{(1)} = \Delta\dot{R}^{(1)}(\Delta\vec{x}) + \Delta_\Delta + \Delta N^{(1)} \cdot \lambda^* + \delta\Delta\dot{L}^{(1)}_\Sigma; \\ p_{\Delta L} & \Delta\hat{\dot{L}}^{(2)} = \Delta\dot{R}^{(2)}(\Delta\vec{x}) + \Delta_\Delta + \Delta N^{(2)} \cdot \lambda^* + \delta\Delta\dot{L}^{(2)}_\Sigma; \\ \cdot & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ p_{\Delta L} & \Delta\hat{\dot{L}}^{(m_L)} = \Delta\dot{R}^{(m)}(\Delta\vec{x}) + \Delta_\Delta + \Delta N^{(m)} \cdot \lambda^* + \delta\Delta\dot{L}^{(m)}_\Sigma; \\ \hline p_{\Delta S} & \Delta\hat{\dot{S}}^{(1)} = \Delta\dot{R}^{(1)}(\Delta\vec{x}) + \Delta_\Delta + \delta\Delta S^{(1)}; \\ p_{\Delta S} & \Delta\hat{\dot{S}}^{(2)} = \Delta\dot{R}^{(2)}(\Delta\vec{x}) + \Delta_\Delta + \delta\Delta S^{(2)}; \\ \cdot & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ p_{\Delta S} & \Delta\hat{\dot{S}}^{(m_S)} = \Delta\dot{R}^{(m)}(\Delta\vec{x}) + \Delta_\Delta + \delta\Delta S^{(m)}, \end{array}\right. \qquad (4)$$

где m – число спутников, для которых имеются приращения наблюдений на текущем интервале Δt; $m \geq 4$; m_L – число приращений фазы; m_S – число приращений кодовых псевдодальностей; $\Delta\vec{x} = \left\|\Delta x, \Delta y, \Delta z\right\|^T$ – искомые приращения координат; $\Delta_\Delta = \Delta\Delta^{GPS}$ – искомое приращение расхождения часов приемника относительно шкалы GPS.

Вектор искомых параметров

$$\vec{\theta} = \left\| \Delta\vec{x}^T, \Delta_\Delta \right\|^T = \left\| \Delta x, \Delta y, \Delta z, \Delta_\Delta \right\|^T. \tag{5}$$

Вектор приращений наблюдений

$$\hat{\vec{\theta}} = \left\| \Delta\hat{L}^{(1)}, \Delta\hat{L}^{(2)}, \ \ldots\ , \Delta\hat{L}^{(m_L)}; \Delta\hat{S}^{(1)}, \Delta\hat{S}^{(2)}, \ \ldots\ , \Delta\hat{S}^{(m_S)} \right\|^T = \left\| \hat{\vec{\theta}}_L^T; \hat{\vec{\theta}}_S^T \right\|^T. \tag{6}$$

Задача заключается в одновременной оценке вектора $\vec{\theta}$ и в обнаружении одновременно одного или нескольких циклических (или полуциклических скачков) ($\Delta N^j = 0$ или $\Delta N^j \neq 0$) в приращениях фазы.

Задача считается решенной с приемлемым качеством, если удается:

- определить элементы вектора $\vec{\theta}$ с точностью не хуже 10 *см* ($\Gamma\Phi \leq 2-4$), невязки между измеренными и вычисляемыми приращениями фазы с точностью не хуже 2,5 – 4 *см*,
- обнаружить, оценить и устранить скачки приращений фазы в наблюдениях одного или нескольких спутников или просто отбраковать решение со скачками, если фазовых наблюдений недостаточно для этого.

В любом случае, должно быть получено решение $\hat{\vec{\theta}}$, даже по приращениям кодовых псевдодальностей, причем во всех случаях должна быть оценена корреляционная матрица погрешностей решения и корреляционная матрица приращений фазовых и кодовых псевдодальностей по корреляционной матрице погрешностей оценок элементов вектора $\hat{\vec{\theta}}$.

Алгоритм оценки $\hat{\vec{\theta}}$ реализуется по следующей стандартной итерационной процедуре метода наименьших квадратов (МНК) для нелинейных систем по схеме линеаризации Ньютона – Гаусса:

$$\begin{cases} \hat{\vec{\theta}}_{(i+1)} = \hat{\vec{\theta}}_{(i)} + \Delta\hat{\vec{\theta}}_{(i+1)}; \ i = 0,1,2, \ldots\text{- номер шага итерационного процесса;} \\ \hat{\vec{\theta}}_{(i+1)} = \left[A_{(i)}^T \cdot W \cdot A_{(i)} \right]^{-1} \cdot A_{(i)}^T \cdot W \cdot \left[\bar{\vec{Q}} - F\left(\hat{\vec{\theta}}_{(i)} \right) \right]. \end{cases} \tag{7}$$

Здесь

$$\underbrace{A_{(i)}}_{(m_L+m_S)\cdot 4} = \left\| \frac{\partial\vec{Q}(\vec{\theta})}{\partial\vec{\theta}} \right\| = \left\| \frac{B_1(i)}{B_S(i)} \right\|;$$

$$\underbrace{W}_{[m_L+m_S]\cdot[m_L+m_S]} = \left\| \begin{array}{ccc|ccc} p_{\Delta L}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & . & 0 & 0 & 0 & 0 \\ 0 & 0 & p_{\Delta L}^2 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & p_{\Delta S}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & . & 0 \\ 0 & 0 & 0 & 0 & 0 & p_{\Delta S}^2 \end{array} \right\| = \left\| \begin{array}{c|c} W_L & 0 \\ \hline 0 & W_S \end{array} \right\|;$$

$$\underbrace{B_{L,S}(i)}_{[m\cdot 4]}=\left\|\begin{array}{cccc}\dfrac{\Delta\hat{x}_{(i)}+\left(\hat{x}_k-\hat{X}_{k+1}^{(1)}\right)}{\hat{R}_{k+1}^{(1)}(i)} & \dfrac{\Delta\hat{y}_{(i)}+\left(\hat{y}_k-\hat{Y}_{k+1}^{(1)}\right)}{\hat{R}_{k+1}^{(1)}(i)} & \dfrac{\Delta\hat{z}_{(i)}+\left(\hat{z}_k-\hat{Z}_{k+1}^{(1)}\right)}{\hat{R}_{k+1}^{(1)}(i)} & 1 \\ \cdots & \cdots & \cdots & \cdots \\ \dfrac{\Delta\hat{x}_{(i)}+\left(\hat{x}_k-\hat{X}_{k+1}^{(m)}\right)}{\hat{R}_{k+1}^{(m)}(i)} & \dfrac{\Delta\hat{y}_{(i)}+\left(\hat{y}_k-\hat{Y}_{k+1}^{(1m)}\right)}{\hat{R}_{k+1}^{(m)}(i)} & \dfrac{\Delta\hat{z}_{(i)}+\left(\hat{z}_k-\hat{Z}_{k+1}^{(m)}\right)}{\hat{R}_{k+1}^{(m)}(i)} & 1\end{array}\right\|;$$

для $m\geq 4$;

или $\begin{cases} m=m_L,\ m_S\geq m_L; \\ m=m_S,\ m_S\geq m_L; \end{cases}$

$$\underbrace{W_L}_{[m\cdot L]}=diag\left\{p_{\Delta L}^2,\ldots,p_{\Delta L}^2\right\};$$

$$\underbrace{W_S}_{[m\cdot S]}=diag\left\{p_{\Delta S}^2,\ldots,p_{\Delta S}^2\right\};$$

$$\hat{R}_{k+1}^{(J)}(i)=\sqrt{\left[\Delta\hat{x}_{(i)}+\left(\hat{x}_k-\hat{X}_{k+1}^{(J)}\right)\right]^2+\left[\Delta\hat{y}_{(i)}+\left(\hat{y}_k-\hat{Y}_{k+1}^{(J)}\right)\right]^2+\left[\Delta\hat{z}_{(i)}+\left(\hat{z}_k-\hat{Z}_{k+1}^{(J)}\right)\right]^2};$$

$$\underbrace{\vec{F}(\hat{\vec{\theta}}_{(i)})}_{[m_L+m_S]}=\left\|\begin{array}{c}\Delta\dot{R}^{(1)}(\Delta\hat{\vec{x}}_{(i)})+\hat{\Delta}_{\Delta}(i) \\ \cdots \\ \underbrace{\Delta\dot{R}^{(m_L)}(\Delta\hat{\vec{x}}_{(i)})+\hat{\Delta}_{\Delta}(i)}_{\vec{F}_L} \\ \Delta\dot{R}^{(1)}(\Delta\hat{\vec{x}}_{(i)})+\hat{\Delta}_{\Delta}(i) \\ \cdots \\ \underbrace{\Delta\dot{R}^{(m_S)}(\Delta\hat{\vec{x}}_{(i)})+\hat{\Delta}_{\Delta}(i)}_{\vec{F}_S}\end{array}\right\|;$$

$$\Delta\dot{R}^{(J)}(\Delta\hat{\vec{x}}_{(i)})=\sqrt{\left[\Delta\hat{x}_{(i)}+\left(\hat{x}_k-\hat{X}_{k+1}^{(J)}\right)\right]^2+\left[\Delta\hat{y}_{(i)}+\left(\hat{y}_k-\hat{Y}_{k+1}^{(J)}\right)\right]^2+\left[\Delta\hat{z}_{(i)}+\left(\hat{z}_k-\hat{Z}_{k+1}^{(J)}\right)\right]^2}-$$

$$-\sqrt{\left(\hat{x}_k-\hat{X}_k^{(J)}\right)^2+\left(\hat{y}_k-\hat{Y}_k^{(J)}\right)^2+\left(\hat{z}_k-\hat{Z}_k^{(J)}\right)^2};$$

$$Q(\vec{\theta})=\left\|Q_L^T\left|Q_S^T\right.\right\|^T.$$

Итерационный процесс будем считать успешно завершенным, если

$\frac{1}{\sqrt{4}}\left[\Delta\hat{\vec{\theta}}_{(i+1)}^T\cdot\Delta\hat{\vec{\theta}}_{(i+1)}\right]^{1/2}\leq\varepsilon$, где ε - заранее заданная остаточная погрешность.

Несколько упростим выражение (7) для оценки поправки $\Delta\hat{\vec{\theta}}_{(i+1)}$:

$$\Delta\hat{\vec{\theta}}_{(i+1)}=\left[B_L^TW_LB_L+B_S^TW_SB_S\right]_{(i+1)}^{-1}\cdot\left[B_L^TW_L\Delta\hat{\vec{Q}}_L+B_S^TW_S\Delta\hat{\vec{Q}}_S\right],\qquad(8)$$

где $\Delta\hat{\vec{Q}}_L = \left\|\Delta Q_L - \vec{F}_L\right\|_{(i+1)}$; $\Delta\hat{\vec{Q}}_S = \left\|\Delta Q_S - \vec{F}_S\right\|_{(i+1)}$.

Корреляционная матрица $K_{\vec{\theta}}$ погрешностей оценки вектора $\hat{\vec{\theta}}$ вычисляется на последнем шаге итераций:

$$\underbrace{K_{\vec{\theta}}}_{[4\cdot 4]} = \left[B_L^T W_L B_L + B_S^T W_S B_S\right]^{-1} \tag{9}$$

Также на последнем шаге вычисляются векторы невязок ΔQ_L и ΔQ_S, элементы которых (для каждого спутника отдельно) могут сравниваться с заранее определенными порогами для получения вывода о наличии или отсутствии скачков в приращениях фазы. Порог для невязок фазы может быть выбран на уровне $\Pi_\Delta = \lambda_1/5$.

Корреляционная матрица оценок приращений фазы $\hat{\vec{F}}_L$, т.е. оценка погрешностей ΔQ_L:

$$\underbrace{K_{\Delta\vec{L}}}_{[m_L \cdot m_L]} = \left\|\frac{\partial\vec{Q}_L(\vec{\theta})}{\partial\vec{\theta}}\right\| \cdot K_{\vec{\theta}} \cdot \left\|\frac{\partial\vec{Q}_L(\vec{\theta})}{\partial\vec{\theta}}\right\|^T = \underbrace{D \cdot K_{\vec{\theta}} \cdot D^T}_{[m_L \cdot 4]}. \tag{10}$$

По диагональным элементам матрицы $K_{\Delta L}$ ($\sigma_{\Delta L_j} = \sqrt{K_{\Delta L}(j,j)}$) можно судить о точности оценки приращений фазы для всех m спутников, даже если часть из них была исключена из обработки в ее процессе. При этом можно рассчитать все приращения фазы и сравнить их с реальными приращениями наблюдений.

Вначале полагаем, что справедлива гипотеза об отсутствии скачков приращений фазы. В этом случае $m_L = m_S$ можно ожидать сходимости итерационного процесса за 1 – 3 итерации. Для минимизации количества итераций целесообразно как можно точнее определить вектор $\hat{\vec{\theta}}_{(0)} = \left\|\Delta\hat{x}_{(0)}, \Delta\hat{y}_{(0)}, \Delta\hat{z}_{(0)}, \Delta_{(0)}\right\|^T$. Для этого необходимо кодовое решение относительно параметров $\Delta\hat{x}(t), \Delta\hat{y}(t), \Delta\hat{z}(t), \hat{\Delta}(t)$ сгладить и затем получить оценки приращений координат и уходов шкалы времени. В этом случае возможно получить решение за одну итерацию.

Если число итераций превышает 5 и процесс не сходится, то можно полагать, что в каком-либо из приращений фазы (или одновременно в нескольких приращениях) присутствуют скачки фазы $\Delta N^{(j)} \neq 0$. Тогда необходимо последовательно проверить гипотезы:

- скачок в одном приращении;
- скачки в более, чем одном приращении.

А) В общем случае должно выполняться условие $m \geq 4$. В этом случае возможно «почувствовать» наличие скачков в фазе несущей. В случае $m = 4$, если даже итерационный процесс сошелся, а анализ невязок (их сравнение с порогом) показал превышение хотя бы в одном наблюдении, необходимо получить решение только по кодовым наблюдениям, вычислить его корреляционную матрицу и пометить решение как грубое.

В) Если $m_L = 5$, и невязки (хотя бы одна) показывают превышение порога, то необходимо перебрать (m_L) вариантов решения с отбрасыванием из вектора наблюдений последовательно одного из m_L значений $\Delta\hat{\vec{L}}^{(j)}$.

B1) Если с исключением одного из наблюдений $\Delta\hat{L}$ точное решение задачи достигается (если фазовые невязки не превышают допустимый порог), то кроме матрицы $K_{\tilde{\theta}}$ рассчитываются значения элементов вектора $\vec{F}_L$ (расчетные значения приращений ΔL из вектора $\Delta\vec{\theta}_L$ и их корреляционная матрица $K_{\Delta\hat{L}}$. Расчет выполняется и для «выпавшего» значения $\Delta\hat{\hat{L}}^{(i)}$. Разность $\Delta\hat{\hat{L}}^{(i)} - \Delta\hat{\hat{L}}^{(i)}_{\text{расч}}$ проверяется на соответствие циклическому или полуциклическому скачку фазы i-го спутника.

Если эта гипотеза подтверждается:

$$\left[\Delta\hat{\hat{L}}^{(i)} - \Delta\hat{\hat{L}}^{(i)}_{(\text{расч})}\right] = \Delta N^{(i)} \cdot \lambda^* + \delta_i,$$

где $|\delta_i| \le 1/5$; $\Delta N^{(i)}$ – целое, то осуществляется коррекция приращения $\Delta\hat{\hat{L}}^{(i)}$ на величину $\Delta N^{(i)} \cdot \lambda^*$, задача решается вновь для полной выборки m_L, фиксируется значение скачка фазы для данного момента времени t_{k+1}.

B2) Если с последовательным исключением наблюдений из выборки $\Delta\hat{\hat{\theta}}$ для оставшихся «четверок» не найдено точного решения, то принимается решение о наличии двух и боле скачков и недостаточности информации для получения точного решения. Оценка приращений координат осуществляется по кодовым наблюдениям.

C) Если $m_L \ge 6$ и решение показывает наличие скачков, то стратегия обработки сводится к следующим действиям:

- поочередно отбрасывается по одному наблюдению, каждый раз анализируются невязки и вычисляется средняя невязка; если в каком-либо варианте отбрасывания наблюдений с оставшимися достигается общее решение (все невязки меньше порога), то отброшенное наблюдение идентифицируется с наличием скачка и оно обрабатывается по приведенной выше схеме с возможность коррекции и повторного использования;

- если все отбрасывания по одному наблюдению не дают эффекта, то отбрасывается то наблюдение, без которого средняя невязка $\sqrt{\delta^2_{(i)}} = \left[\frac{1}{m_L} \cdot \Delta\hat{\hat{\theta}}_L^T \cdot \Delta\hat{\hat{\theta}}_L\right]^{1/2}$ – наименьшая; далее действия повторяются с укороченной выборкой на одно наблюдение;

- если отбрасывание и второго наблюдения не дало эффекта, то по критерию влияния отброшенного наблюдения на среднее значение невязки бракуется теперь и второе наблюдение;

- процесс повторяется циклически с еще раз укороченной выборкой;

- если не найдено приемлемого решения и далее по фазе, то окончательно формируется кодовое решение.

В принципе, можно остановиться (для ускорения и упрощения решения) на усеченном варианте стратегии (с проверкой за один подход):

$$\begin{cases} HipA: \text{ в наблюдениях 1 скачок} \rightarrow \text{ фазовое решение;} \\ HipB: \text{ в наблюдениях более 1 скачка} \rightarrow \text{ кодовое решение.} \end{cases}$$

В случае выполнения обработки наблюдений фазы L_1 базовых станций (приемников) в статическом режиме априори известно, что изменение координат:

$$\Delta\vec{x} = \begin{vmatrix} \Delta x \\ \Delta y \\ \Delta z \end{vmatrix} \equiv \vec{0}.$$

Это свойство можно использовать для контроля качества починки скачков фазы L_1 наблюдений неподвижного ГНСС-приемника.

Для этого случая система уравнений наблюдений (4) с учетом предварительного расчета и исключения из наблюдений изменения дальностей $\Delta\dot{R}^{(j)}(\Delta\vec{x},\vec{0})$, обусловленных движением спутников, ухода часов спутников, изменения тропосферной и ионосферной погрешностей (с использованием модели) может быть трансформирована в систему (11):

$$\begin{cases} \Delta\hat{\dot{L}}^{*(j)} = \Delta_\Delta + \Delta N^{(j)} \cdot \lambda^* + \delta\Delta\dot{L}^{(j)}; \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \Delta\hat{\dot{S}}^{*(j)} = \Delta_\Delta + \delta\Delta\dot{S}^{(j)}, \end{cases} \tag{11}$$

где $\Delta\hat{\dot{L}}^{*(j)}$, $\Delta\hat{\dot{S}}^{*(j)}$ – наблюдения приращений фазовых и кодовых псевдодальностей, скорректированных на известную часть приращений.

Система уравнений (11) является линейной, и задача сводится к обнаружению и оценке скачков фазы $\Delta N^{(j)}$. Оценка ухода часов, в данном случае не является обязательным действием, хотя, если устранить фазовые скачки на участках и оценить точные значения приращений уходов часов, то можно очень точно исключить их из наблюдений кода и фазы.

Теперь рассмотрим алгоритм обнаружения скачков фазы.

Если взять наблюдения только $\Delta\hat{\dot{L}}^{*}$ (фазовые наблюдения), то для обнаружения скачков можно использовать уже апробированный алгоритм анализа разностей приращений фазы между спутниками, т.е.

$$\nabla\Delta\hat{\dot{L}}^{*(i,1)} = \Delta\hat{\dot{L}}^{*(i)} - \Delta\hat{\dot{L}}^{*(1)}, \ (i = 2,\ldots,m).$$

В этих разностях изменения уходов часов приемника Δ_Δ компенсируются и, если какая-либо разность $\nabla\Delta\hat{\dot{L}}^{*(i,1)}$ превышает допустимый порог (относительно шумов), равный примерно $\lambda_1/5$, то фиксируется это превышение (скачок) и оценивается близость скачка к целому количеству циклов λ_1 (или к полному количеству полуциклов $\lambda_1/2$). Если проанализировать все разности относительно одного и того же спутника и применить принцип – «наблюдения трех спутников одновременно не могут иметь (по вероятности) одинаковые скачки фазовых наблюдений», – то можно идентифицировать либо два и более «нулевых» скачков либо три и более одинаковых скачков разностей приращения фазы.

Обнаруженные и оцененные скачки устраняются из разностей приращений фазы и даже может быть получена оптимальная МНК-оценка значения Δ_Δ согласно выражению (8) с учетом линейности задачи и только одного оцениваемого параметра:

$$\hat{\Delta}_\Delta = \frac{\sum_{j=1}^{m}\left[p^2_{\Delta L} \cdot \Delta\hat{\dot{L}}^{*(j)} + p^2_{\Delta S} \cdot \Delta\hat{\dot{S}}^{*(j)} \right]}{m \cdot (p^2_{\Delta L} + p^2_{\Delta L})}. \tag{12}$$

Рассмотрим также случай обработки наблюдений фазы L_1 «роверного» приемника в кинематическом режиме съемки с использованием предварительно обработанных наблюдений базовой станции.

В данном случае полагаем, что при обработке наблюдений «роверного» приемника формируются одинарные разности фазовых (и кодовых) наблюдений между «роверным» приемником и базовой станцией. Преимущества дифференциального режима в этом случае очевидны (см. выше) – исключаются погрешности часов спутника, уменьшаются эфемеридные и атмосферные погрешности.

Тогда приращения по времени (3) могут быть записаны в виде

$$\begin{cases} \Delta\hat{L}^j = \left[\Delta\dot{R}^j_{\text{П}} - \Delta\dot{R}^j_{\text{реф}}\right] + \left[\Delta\Delta^{\text{GPS}}_{\text{П}} - \Delta\Delta^{\text{GPS}}_{\text{реф}}\right] + \Delta N^j_{\text{П}} \cdot \lambda^* + \delta\Delta^j_{\text{эф}} + \delta\Delta T^j + \\ \quad + \delta\Delta I^j + \left[\delta\Delta L^j_{\text{П}\,mf} - \delta\Delta L^j_{\text{реф}\,mf}\right]; \\ \Delta\hat{S}^j = \left[\Delta\dot{R}^j_{\text{П}} - \Delta\dot{R}^j_{\text{реф}}\right] + \left[\Delta\Delta^{\text{GPS}}_{\text{П}} - \Delta\Delta^{\text{GPS}}_{\text{реф}}\right] + \delta\Delta^j_{\text{эф}} + \delta\Delta T^j + \\ \quad + \delta\Delta I^j + \left[\delta\Delta S^j_{\text{П}\,m_f} - \delta\Delta S^j_{\text{реф}\,m_f}\right], \end{cases} \tag{13}$$

где при сохранении общих обозначений индексы «П» и «реф» означают – «потребитель» (ровер) и референцная (базовая) станция соответственно; $\delta\Delta_{\text{эф}}, \delta\Delta T, \delta\Delta I$ – разностные (для текущей базы между «ровером» и базовой станцией) значения эфемеридных, тропосферных и ионосферных погрешностей соответственно;

$$\Delta\Delta^{\text{GPS}}_{\text{П}} - \Delta\Delta^{\text{GPS}}_{\text{реф}} = \Delta\Delta_{\text{П-реф}};$$

$$\Delta\dot{R}^j_{\text{реф}} = \sqrt{\left(\hat{x}^{\text{реф}} - \hat{X}^j_{k+1}\right)^2 + \left(\hat{y}^{\text{реф}} - \hat{Y}^j_{k+1}\right)^2 + \left(\hat{z}^{\text{реф}} - \hat{Z}^j_{k+1}\right)^2} - \sqrt{\left(\hat{x}^{\text{реф}} - \hat{X}^j_{k}\right)^2 + \left(\hat{y}^{\text{реф}} - \hat{Y}^j_{k}\right)^2 + \left(\hat{z}^{\text{реф}} - \hat{Z}^j_{k}\right)^2}.$$

После ввода коррекций и учета $\Delta\dot{R}^j_{\text{реф}}$ получим систему уравнений – аналог системы (4) с той разницей, что вместо параметра Δ_Δ оценивается параметр $\Delta\Delta_{\text{П}-\text{реф}}$ (в дифференциальных наблюдениях увеличиваются шумовые и многолучевые составляющие погрешностей, но значительно уменьшаются медленно меняющиеся погрешности).

Далее решение системы уравнений выполняется по алгоритму, описанному ранее.

Целесообразно также рассмотреть возможность использования двухчастотных фазовых наблюдений для коррекции ионосферной составляющей погрешностей приращения фазы наблюдений диапазона L_1. Поскольку при обработке приращений фазы L_1 желательно как можно точнее учесть ионосферную составляющую для больших интервалов обработки разности фаз по времени, то при наличии двухчастотных наблюдений «роверного» приемника или станции (как в абсолютном, так и в дифференциальном режимах) можно практически полностью исключить ионосферную составляющую приращений фазы следующим образом:

- формируется для заданного участка наблюдений j-го спутника разность

$$\left(L_1 - L_2\right) = \left(\gamma^2 - 1\right) \cdot I + \underbrace{\left(\Delta N \cdot \lambda_1 - \Delta N \cdot \lambda_2\right)}_{const} + c + \left(\delta L_1 - \delta L_2\right);$$

эта разность может содержать не только неизвестные константы, но и «слипы» в наблюдениях фазы обеих частот;

- линейная комбинация (L_1-L_2) (в ней исключена геометрия, эфемериды, часы и тропосфера) подвергается (с использованием процедуры оценки «слипов») обработке, в ходе которой можно получить сглаженное значение «безгеометрической» функции (L_1-L_2) с точностью до неизвестной константы, причем с интерполяцией и экстраполяцией на моменты пропущенных наблюдений. Эта кривая является гладкой и не содержит выбросов из-за «слипов», поэтому оценки ее изменения (приращения) с высокой точностью соответствуют изменениям ионосферной погрешности;

- простое масштабирование дает возможность сформировать и исключить ионосферные коррекции в приращении фазы L_1.

Такой подход дает возможность увеличить интервал формирования и обработки приращений фазы L_1.

Рассмотрим теперь использование оценок приращений фазы L_1 на смежных интервалах для коррекции (компенсации) динамики «роверного» приемника и сформулируем рекомендации по использованию полученных оценок.

1) В результате обработки по смежным моментам времени t_{k+1} и t_k имеем оценки приращений: $\Delta\hat{x}_{k+1}, \Delta\hat{y}_{k+1}, \Delta\hat{z}_{k+1}, \hat{\Delta}_{\Delta_{k+1}}$. Кроме того, для этих оценок имеем оценки корреляционных матриц. Если в каких-либо точках отсутствуют точные оценки (по фазовым наблюдениям), либо были получены грубые оценки (по кодовым наблюдениям), то целесообразно прогладить оценки $\Delta x.\Delta y, \Delta z, \Delta_{\Delta_{k+1}}$, чтобы интерполировать оценки на отсутствующие точки, либо на точки (моменты), где были получены грубые оценки. В итоге имеем точные значения приращений.

2) Теперь можно получить текущее изменение координат «ровера» относительно начальной точки всего сеанса измерений, т. е. координаты и время приемника, с точностью до начальной константы:

$$\Delta\hat{x}(t_k) = \sum_{L=0}^{k-1} \Delta\hat{x}_{L+1};\ \ \Delta\hat{y}(t_k) = \sum_{L=0}^{k-1} \Delta\hat{y}_{L+1};\ \ \Delta\hat{z}(t_{k+1}) = \sum_{L=0}^{k-1} \Delta\hat{z}_{L+1};$$

$k = 1, K;\ (K - \text{число эпох сеанса});$

$$\hat{\Delta}(t_k) = \sum_{L=0}^{k-1} \hat{\Delta}_{\Delta} \hat{R}_{L+1};\ \ \Delta\hat{x}_1 = \Delta\hat{y}_1 = \hat{\Delta}\hat{z}_1 \equiv 0.$$

3) Чтобы получить абсолютные значения текущих координат и уходов шкалы времени «роверного» приемника, необходимо по известной методике:

- определить разность кодовых оценок $\hat{x}, \hat{y}, \hat{z}, \hat{\Delta}$ и фазовых оценок $\hat{x}(t_k), \hat{y}(t_k), \hat{z}(t_k), \hat{\Delta}(t_k)$, полученных ранее;

- сгладить полученные разности, используя, например, процедуры сплайн–аппроксимации;

- сложить медленно меняющуюся сглаживающую функцию и фазовые оценки координат и времени.

4) Теперь полученные «бесшумовые» значения координат и времени можно использовать для расчета фазы псевдодальностей и устранения динамики в наблюдениях по приведенным выше соотношениям.

5) Целесообразно также рассмотреть задачу контроля качества починки фазовых скачков с использованием измененного метода обработки приращений. При изложенной обработке приращений (главным образом для темпа наблюдений от 1 до 1/5 *Гц*) осуществляется обнаружение (и, по возможности, устранение) скачков на смежных интервалах (1 – 5 *с*).

Здесь возникает задача контроля скачков на больших допустимых интервалах (1 – 15 *с* для автономного режима и до 30 *с* для дифференциального) тех или иных спутников. Для реализации контроля необходимо:

- определить приращения координат и времени на участках пропусков наблюдений по описанной выше схеме;
- рассчитать с учетом всех поправок приращения фазы для анализируемых участков;
- сравнить измеренные и расчетные приращения фаз, проанализировать невязки, сравнив с порогом.

Таким образом, в данном подразделе детально изложены предложенные алгоритмы обнаружения и оценки фазовых скачков в кинематических ГНСС наблюдениях и приведены практические рекомендации по их применению и ограничения.

Выполненные многочисленные экспериментальные исследования с использованием реальных наблюдений двухчастотных ГНСС приемников, выполненных в районах Житомира, Винницы и Одессы в процессе аэрофотосъемок на борту самолета и вертолета и на базовых станциях (данные представлены коммерческим предприятием «Магеллан», Киев), показали, что расчетные оценки допустимых интервалов взятия приращений соответствуют реальным значениям. Реализация и тестирование предложенных алгоритмов обработки наблюдений с высокой динамикой в программном комплексе «OCTAVA_PPA» доказали справедливость теоретических исследований и возможность выполнения надежной предварительной обработки фазовых наблюдений, полученных в кинематическом режиме съемки.

Обработка кинематических фазовых наблюдений. Альтернативный метод

Как упоминалось выше, расширить возможности и повысить надежность устранения фазовых скачков одночастотных и двухчастотных кинематических наблюдений в сложных условиях позволяет применение другого, альтернативного, более эффективного и надёжного метода восстановления непрерывности фазовых наблюдений для условий полной или частичной кратковременной потери наблюдений [15, 16], который основан на использовании приращений наблюдений по времени и специальной процедуры совместного МНК–оценивания совокупности континуальных (приращений координат и расхождений часов) и дискретных параметров (циклических фазовых скачков). Здесь приводится краткое описание указанного метода решения поставленной задачи и главные результаты его тестирования с использованием реальных ГНСС наблюдений.

В ходе исследований был создан новый эффективный и надёжный метод восстановления непрерывности фазовых двухчастотных и одночастотных наблюдений (обнаружения, оценки и устранения дискретных фазовых скачков) для условий полной или частичной кратковременной потери двухчастотных наблюдений [15, 16]. Исследования проводились для случаев полной потери наблюдений всех спутников текущего созвездия на интервале 5 – 30 *с* и потери наблюдений части спутников текущего созвездия на заданном интервале. Подход, изложенный в [15], позволяет восстанавливать непрерывность фазовых наблюдений даже при полной потери слежения за спутниками на интервалах до 20 *с*. При этом, в качестве начальной априорной информации используются исходные кодовые наблюдения без применения специальных процедур их сглаживания. Использование фазовых наблюдений разностной частоты (Wide Lane) в качестве промежуточной шкалы позволяет обеспечить высокую надёжность устранения фазовых циклических скачков на частоте L_1. Однако, с учетом широкого практического использования одночастотной аппаратуры, актуальным и важным для практики является использование предложенного подхода для решения задачи обнаружения, оценки и устранения циклических фазовых скачков одночастотных кинематических ГНСС-наблюдений.

Для выделения фазовых скачков и уменьшения остаточных погрешностей тропосферных и ионосферных задержек, для обработки используются приращения по времени кодовых и фазовых наблюдений (без потери общности будем здесь рассматривать GPS наблюдения).

Это позволяет устранить инструментальные задержки в передающих трактах спутников и приёмном тракте приёмника, а также случайную начальную фазу опорного генератора на спутниках:

$$\begin{cases} \Delta\hat{\mathbf{S}}^{j}_{1,2i} = \Delta\mathbf{F}^{j} + k_{1,2}\cdot\Delta\mathbf{I}^{j} + \delta\Delta\mathbf{S}^{j}_{1,2i} \\ \Delta\hat{\mathbf{L}}^{j}_{1,2i} = \Delta\mathbf{F}^{j} - k_{1,2}\cdot\Delta\mathbf{I}^{j} - \Delta\mathbf{N}^{j}_{1,2}\cdot\lambda_{1,2} + \Delta\varphi_{0}\cdot\lambda_{1,2} + \delta\Delta\mathbf{L}^{j}_{1,2i} \end{cases}, \quad (14)$$

где $\Delta\hat{\mathbf{S}}^{j}_{1,2i}$ – приращение кодовой псевдодальности между соседними временными эпохами (между j-м спутником и i-м приёмником) на 1-й и на 2-й частотах соответственно;

$\Delta\hat{\mathbf{L}}^{j}_{1,2i}$ – приращение фазовой псевдодальности между соседними временными эпохами (между j-м спутником и i-м приёмником) на 1-й и на 2-й частотах соответственно;

$$\Delta\mathbf{F}^{j} = \Delta\mathbf{R}^{j}_{i} + [\Delta\Delta^{GPS}_{i} - \Delta\Delta^{j,GPS}] + \Delta\mathbf{T}^{j}_{i};$$

$\Delta\mathbf{R}^{j}_{i}$ – приращение геометрического расстояния между i-м приёмником и j-м спутником;

$\Delta\Delta^{GPS}_{i}$ – приращение отклонений шкалы времени часов приёмника от системного времени;

$\Delta\Delta^{j,GPS}$ – приращение отклонений шкалы времени часов спутника от системного времени;

$\Delta\mathbf{T}^{j}_{i}$ – приращение тропосферной задержки сигнала;

$\Delta\mathbf{I}^{j}$ – приращение ионосферной задержки сигнала;

$k_1=1$; $k_2=\gamma^2$ – ионосферная константа;

$\delta\Delta\mathbf{S}^{j}_{1,2i}$, $\delta\Delta\mathbf{L}^{j}_{1,2i}$ – шумы и многолучевость кодовых и фазовых наблюдений соответственно;

$\Delta\mathbf{N}^{j}_{1,2}$ – приращения целочисленной фазовой неоднозначности (величина фазового скачка) – $\Delta\mathbf{N}^{j}_{1,2}=0$, если скачок отсутствует;

$\lambda_{1,2}$ – длина волны на 1-й и на 2-й частотах соответственно.

Полагаем, что на всем интервале наблюдений приёмник не выключался, поэтому $\Delta\varphi_0$ равно нулю (приращение случайной начальной фазы опорного генератора приёмника).

Также полагаем, что в представленные приращения наблюдений введены модельные (или дифференциальные и модельные) коррекции тропосферной и ионосферной составляющих с использованием текущего решения навигационной задачи (НЗ). Эфемеридная составляющая будет пренебрежимо мала, так что на интервале $\Delta t = (t_{k+1} - t_k) = 10 - 15$ *с* максимальные ($\delta=3\sigma$) медленно меняющееся погрешности фазы ($\hat{L}_1, \hat{L}_2$) не превысят $1 - 2$ *см* вместе с суммарными флуктуационными и многолучевыми погрешностями. Составляющей $\Delta\Delta^{j,GPS}$ на рассматриваемом интервале $\Delta t=10 - 15$ *с* можно пренебречь. Таким образом, суммарные погрешности приращений фазовых и кодовых наблюдений можно оценить как $\delta_{L1,L2}=1 - 2$ *см*, $\delta_{S1,S2}=1 - 2$ *м*.

Положим, что имеем кодовое решение НЗ в моменты t_{k+1}, t_k. Используя линейное разложение $\Delta\mathbf{F}^j(t_{k+1}, t_k)$ в ряд Тейлора (линеаризация) и введя в наблюдения тропосферные и ионосферные коррекции, получим

$$\begin{cases} \Delta\hat{\mathbf{S}}^j_{1,2} - \Delta\hat{\mathbf{F}}^j = \Delta\hat{\mathbf{S}}^j_{1,2} = \left\|\hat{\vec{a}}^j_{k+1}\right\|^T \cdot \Delta\vec{x} + \Delta_\Delta + \delta\Delta\overset{\circ}{\mathbf{S}}{}^j_{1,2} \\ \Delta\hat{\mathbf{L}}^j_{1,2} - \Delta\hat{\mathbf{F}}^j = \Delta\hat{\mathbf{L}}^j_{1,2} = \left\|\hat{\vec{a}}^j_{k+1}\right\|^T \cdot \Delta\vec{x} + \Delta_\Delta - \Delta\mathbf{N}^j_{1,2} \cdot \lambda_{1,2} + \delta\Delta\overset{\circ}{\mathbf{L}}{}^j_{1,2} \end{cases}, \quad (15)$$

где $\left\|\hat{\vec{a}}^j_{k+1}\right\|$ – вектор направляющих косинусов j-го спутника относительно приёмника; $\Delta\vec{x}$ – вектор поправок к приращениям координат; Δ_Δ – поправка к приращению времени.

В приведенных системах уравнений (14), (15) для фазовых наблюдений опущены составляющие погрешностей, обусловленные влиянием так называемого «wind-up»-эффекта, который является дополнительным фазовым набегом из-за изменения взаимной ориентации (вращения) антенн спутника и приёмника. Фазовыми набегами, которые обусловлены вращением спутников ГНСС вокруг своей оси в процессе ориентирования на Солнце, в приращениях по времени на интервалах в несколько десятков секунд можно пренебречь. Фазовыми же набегами («wind-up»-эффект), обусловленными вращением приемной антенны, пренебрегать нельзя, так как величина этого эффекта на указанных интервалах наблюдений может составлять от долей цикла до нескольких циклов, в зависимости от скорости вращения антенны. Но, поскольку этот эффект является одинаковым для всех спутников текущего созвездия, то его можно объединить и оценивать совместно с уходом шкалы времени приёмника относительно шкалы времен ГНСС. Уравнения (15), в общем случае, также включают неизвестные целочисленные неоднозначность $\Delta\mathbf{N}^j$. Неопределенность выражается в возможности одновременного варьирования величины $\Delta\mathbf{N}^j$ и Δ_Δ при сохранении одной и той же величины их разности. Ввиду этого целесообразно трансформировать (15) в эквивалентную систему уравнений, в которой будут явно выражены двойные разности целочисленных переменных $\nabla\Delta\mathbf{N}^j \cdot \lambda$ без явного формирования разностей уравнений. Соответствующие преобразования системы уравнений (15) и построение алгоритма оценки циклических фазовых скачков показаны в работе [15].

Задача сводится к следующим действиям:

- после преобразований (декорреляции и приведения к компактной форме) исходных уравнений согласно [15] и находится МНК-оценка целочисленного вектора неоднозначностей (искомых фазовых скачков) как континуальных переменных (float-оценка);
- полученные оценки неоднозначностей округляются до ближайшего целого и результат округления принимается за начальное приближение искомого вектора неоднозначностей;
- исходя из оценки корреляционной матрицы float-оценки неоднозначностей, определяется диапазон целочисленного перебора элементов вектора неоднозначностей в окрестности float-оценки;
- осуществляется поиск оптимального целочисленного решения путем минимизация целевой функции (суммы квадратов невязок преобразованного вектора наблюдений и подбираемой функции – см. [15]).

Как и в случае обработки двухчастотных фазовых наблюдений и в случае обработки одночастотных наблюдений [16], существуют два возможных варианта разрыва фазовых наблюдений. Во-первых, возможен предельный случай, когда между моментами времени t_{k+1}, t_k имеется разрыв всех наблюдений и существует априорная неопределенность в отношении возможных скачков фазы всего текущего созвездия спутников. Во-вторых, наиболее вероятна ситуация, когда возникают пропуски наблюдений фазы из-за временных потерь слежения (а значит, возможно и возникновение фазовых скачков) для одного или нескольких

спутников, но не всех спутников текущего созвездия. Особенности обработки одночастотных наблюдений заключаются в том, что не существует возможности сформировать промежуточную измерительную шкалу (Wide Lane) для повышения надёжности решения задачи. Поэтому процедура восстановления непрерывности одночастотных наблюдений, с одной стороны, упрощается, а с другой – не имеет возможности использовать высокоточную дополнительную информацию (наблюдения Wide Lane). Очевидно, что следует ожидать снижения надежности обнаружения и оценки целочисленных скачков L_1 наблюдений по сравнению с двухчастотным случаем.

В случае полного разрыва наблюдений на первом этапе обработки определяются двойные разности фазовых неоднозначностей на первой частоте ($\nabla\Delta\vec{\mathbf{N}}_{L1}$). А на втором, после оценки $\nabla\Delta\hat{\vec{\mathbf{N}}}_{L1}$ и введения коррекций в фазовые наблюдения $\Delta\hat{\mathbf{L}}$, определяются одинарные разности циклических фазовых скачков $\Delta\vec{\mathbf{N}}_{L1}$. В случае частичного разрыва наблюдений, на первом же этапе обработки возможна оценка одинарных разностей циклических фазовых скачков $\Delta\vec{\mathbf{N}}_{L1}$. Для повышения надёжности решения применяются дополнительные подходы, которые детально описаны в работах [15, 16].

В ходе статистического моделирования и тестирования предложенного альтернативного метода обнаружения и оценки фазовых циклических скачков с использованием реальных данных были получены следующие выводы и практические рекомендации.

Двухчастотный случай

1) Предложенный метод позволяет получить надежное решение (с вероятностью $p \geq 0{,}99$) как при частичной, так и при полной потере наблюдений всех спутников текущего созвездия, даже при использовании одной группы приращения между двумя соседними эпохами. При этом рекомендуется использовать одновременно наблюдения шести и более спутников. Надежное решение можно получить на интервалах пропусков наблюдений от 1 до 10 *с* при одновременном наблюдении шести-семи спутников и на интервалах пропусков до 20 *с* при одновременном наблюдении не менее восьми спутников.

2) При использовании нескольких групп приращений (накопление данных), надёжность решения значительно повышается. Использование группы из пяти приращений позволяет увеличить допустимый уровень (СКП) шума фазовых наблюдений в 2 – 3 раза для случая полной потери наблюдений, и в 1,5 – 2 раза для случая частичной потери наблюдений.

3) Использование двухчастотной линейной комбинации Wide Lane (фаза разностной частоты) позволяет также существенно уменьшить неопределенность области перебора целочисленных фазовых неоднозначностей. Это дает возможность уменьшить объём вычислительных операций при выполнении перебора. Использование малых объемов выборок наблюдений, а также уменьшение объёма вычислительных операций из-за уменьшения области неопределенности при использовании комбинации Wide Lane, позволяет выполнять обработку и в реальном времени.

Одночастотный случай

1) При использовании кодовых наблюдений без предварительного их сглаживания предложенный метод имеет ограничения его применения, так как позволяет получать надежные оценки циклических фазовых скачков только в случае потери наблюдений за частью спутников текущего созвездия, когда число спутников с непрерывными наблюдениями не менее четырех. При использовании нескольких групп приращений (накопление данных) надёжность решения повышается, но и в этом случае без использования сглаженных кодовых наблюдений удовлетворительная надежность решения поставленной задачи не достигается, особенно в случае полной кратковременной потери слежения за спутниками. Использование дифференциального режима также является способом повышения качества решения рас-

сматриваемой задачи, поскольку позволяет уменьшить медленно меняющиеся составляющие погрешностей приращений фазовых и кодовых наблюдений. Однако и этот путь не приводит к ожидаемому результату.

2) Определяющим фактором в достижении высокой надёжности решения поставленной задачи является использование сглаженного кодово-фазового координатного решения, что может быть рекомендовано и для случая обработки двухчастотных ГНСС-наблюдений. Применение сглаженных кодовых наблюдений, дифференциального режима и обработки групп приращений позволяет наиболее надёжно восстанавливать непрерывность фазовых наблюдений при полной либо частичной потере слежения за спутниками на интервалах до 30 *с* при использовании наблюдений шести и более спутников.

Рассмотренный альтернативный метод и реализующие его алгоритмы решения задачи восстановления непрерывности кинематических спутниковых наблюдений практически без изменений могут быть применены к наблюдениям не только спутников GPS, но и других ГНСС: ГЛОНАСС, Galileo, Compass.

Заключение

1. Рассмотрена задача определения потерь счета циклов в одночастотных и двухчастотных фазовых ГНСС-наблюдениях и их исправления (восстановления непрерывности). Описаны результаты разработки и тестирования новых эффективных методов и алгоритмов обнаружения, оценки и исправления фазовых скачков как статических, так и кинематических ГНСС-наблюдений.

2. Рассмотрены два отличающихся метода обработки наблюдений. Первый из них реализует интервальную итерационную процедуру обнаружения и оценки (фиксации) скачков одинарных разностей фаз между парами спутников с последующей их идентификацией на отдельных трассах «спутники – приемник». Этот метод, как показали многочисленные экспериментальные исследованиям, оказался особенно эффективен для обработки статических наблюдений, а его модификация позволила решить задачу восстановления непрерывности и кинематических фазовых наблюдений.

3. В ходе обработки разреженных наблюдений (темп данных 1/5 – 1/30 *Гц*) перманентных референцных IGS/EPN станций предложено использовать все возможные двойные разности наблюдений с целью исключить вариации часов спутников в увеличенных интервалах между эпохами и повышения надежности обнаружения и точности оценки скачков. В этом случае для повышения надежности обработки и идентификации скачков на каждой отдельной трассе «спутник – приемник» в структуру наблюдений станции сети целесообразно включить предварительно обработанные наблюдения станций с темпом данных 1 – 1/5 *Гц*.

В ходе обработки кинематических наблюдений с высокой динамикой перед использованием основного метода обнаружения, оценки и устранения фазовых скачков предложено применять предварительную процедуру обработки, которая позволяет оценить и компенсировать динамику наблюдений на трассах «спутники – приемник». С этой целью выполняется точная оценка приращений координат и уходов часов приемника с использованием приращений соседних эпох наблюдений фаз с параллельной отбраковкой возможных фазовых выбросов и скачков.

4. Расширить возможности и повысить надежность устранения фазовых скачков одночастотных и двухчастотных кинематических наблюдений в сложных условиях (разрывы и пропуски наблюдений) позволяет применение другого альтернативного метода решения поставленной задачи. Метод основан на использовании приращений наблюдений по времени и специальной процедуры совместного МНК-оценивания совокупности континуальных (приращений координат и расхождений часов) и дискретных параметров (циклических фазовых скачков). Оценивание циклических фазовых скачков осуществляется прямым перебором в предварительно определённой области поиска. Выполненное тестирование метода с ис-

пользованием реальных ГНСС-наблюдений показало достижение надёжного устранения фазовых скачков при кратковременной полной либо частичной потере наблюдений длительностью до 20 – 30 *с*.

Список литературы: 1. *Гофманн-Веллінгоф, Б.* Глобальна система визначення місцеположення (GPS). Теорія та практика: пер. з англ. / Б. Гофманн-Веллінгоф, Г. Ліхтенеггер, Д. Коллінз ; пер. з англ за ред. Я. С. Яцківа. – Київ : Наук. думка, 1996. – 380 с. 2. *Rizos Chris.* Principles and practice of GPS surveying // School of engineering. The University of New South Wales, Australia, 1999. – 555 p. 3. *Leick, A.* GPS Satellite Surveying, 3-rd ed., John Wiley – New York, 2003. – 464 pp. 4. *Поваляев, А.А.* Спутниковые радионавигационные системы. Время, показания часов, формирование измерений и определение относительных координат. – М. : Радиотехника, 2008. – 328 с. 5. *Bock, Ye., Gourevitch, S.A., Counselman, C.C., King, R.W. and Abbot, R.I.* Interferometric analysis of GPS observations // Manuscripta Geodaetica, Springer-Verlag. – 1986. – №11. – P. 282-288. 6. *Bisnath SB* (2000) Efficient, automated cycle-slip correction of dual-frequency kinematic GPS data, Proceedings of ION GPS 2000, Salt Lake City, Utah, 19-22 Sept., 145-154. 7. *Bisnath SB, Kim D,* and Langley RB. A new approach to an old problem: carrier-phase cycle slips, GPS World, Vol. 12, No. 5, May, 46-51. 8. *Kim, D.*, Langley RB (2001). Instantaneous real time cycle-slip correction of dual-frequency GPS data,, Proceedings of the International Symposium on Kinematic Systems in Geodesy, Geomatics and Navigation, Banff, Alberta, Canada, 5-8 June, 255-264. 9. *Roberts, G.W., Meng, X., Dodson, A.H.* (2002). Using adaptive filtering to defect multipath and cycle slips in GPS/Accelerometer bridge deflection monitoring data, FIG XXII International Congress, TS6.2 Engineering Surveys for Construction Works and Structural Engineering, Washington, D.C. USA, 19-26 April. 10. *Zhalilo, A.A.* Carrier-Phase Cycle-Slip detection and repair of Dual-Frequency GPS data – New Technique using Correlation Filtering Principle // Proceedings of the 10th Saint Petersburg International Conference on Integrated Navigation Systems. – St. Petersburg, Russia, 26 – 28 May, 2003. – P. 273 – 276. 11. *Zhalilo, A., Sadanova, N.* Pre-Processing & Analysis software "OCTAVA_PPA": concept, possibilities and features, initial test results, The Proceedings of the 2004 International Symposium on GNSS/GPS, Sydney, Australia, 6–8 December 2004, CD, 17 pp. 12. *Шелковенков, Д.А., Жалило, А.А.* Возможности повышения надежности обнаружения, оценки и коррекции циклических фазовых скачков двухчастотных наблюдений перманентных GPS-станций в режиме постсеансной обработки // Сб. тр. 2-го Междунар. радиоэлектронного Форума (МРФ-2005) «Прикладная радиоэлектроника. Состояние и перспективы развития». Междунар. конф. по системам локации и навигации (МКЛСН-2005), г. Харьков: ХНУРЭ, 19-23 сентября, 2005 г., Т. 2, С. 544-547. 13. *Жалило, А.А.* Обнаружение и устранение фазовых циклических скачков одночастотных и двухчастотных GPS/GNSS наблюдений – новые универсальный метод и алгоритмы // Тр. XIV-й Санкт-Петербургской междунар. конф. по интегрированным навигационным системам. – Санкт-Петербург, Россия, 28-30 мая 2007 г. – С. 293-302. 14. *Zhalilo, A., Shelkovenkov, D.* Features and service performance of multifunctional software toolkit "OCTAVA" for processing and analysis of GPS/GNSS observations, GEOS 2007 Conference Proceedings, Prague, Czech Republic, 1st – 2nd March 2007, pp. 102-110. 15. *Жалило, А.А., Дицкий, И.В.* Новый эффективный метод устранения циклических фазовых скачков двухчастотных кинематических ГНСС–наблюдений // Известия вузов. Радиоэлектроника. – 2011. – Т. 54, №8. – С. 18-28. 16. *Жалило, А.А., Дицкий, И.В.* Устранение циклических фазовых скачков одночастотных кинематических ГНСС-наблюдений // Известия вузов. Радиоэлектроника. – 2012. – Т.55, №8. – С. 40-52.

Харьковский национальный
университет радиоэлектроники

Поступила в редколлегию 23.09.2012

БЛОЧНЫЕ СИММЕТРИЧНЫЕ ШИФРЫ
BLOCK AND SYMMETRIC CYPHERS

УДК 621. 3.06

О роли схем разворачивания ключей в атаках на итеративные шифры / *В.И. Долгов, А.А. Настенко* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 7 – 15.

Показано, что как и для других известных шифров показатели стойкости шифра DES к атакам линейного и дифференциального криптоанализа от схем разворачивания ключей практически не зависят. При любых схемах разворачивания ключей шифр DES после 16 циклов приходит к показателям случайной подстановки соответствующей степени. Отмечается, что утверждение о том, что для любых ключевых графиков и большинства шифров, отнесенных к марковским, распределения вероятностей дифференциалов и линейных оболочек сходятся к равномерному распределению после некоторого количества циклов, является неверным.

Табл. 4. Библиогр.: 18 назв.

УДК 621. 3.06

Про роль схем розгортання ключів у атаках на ітеративні шифри / *В.И. Долгов, А.А. Настенко* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 7 – 15.

Показано, що як і для інших відомих шифрів показники стійкості шифру DES до атак лінійного та диференційного криптоаналізу від схем розгортання ключів практично не залежать. При будь-яких схемах розгортання ключів шифр DES після 16 циклів приходить до показника випадкової підстановки відповідного ступеня. Відмічається, що твердження про те, що для будь-яких ключових графіків і більшості шифрів, віднесених до марківських, розподілення ймовірностей диференціалів і лінійних оболонок сходиться до рівномірного розподілу нию після деякої кількості циклів, є невірним.

Таб. 4. Бібліогр. :18 назв.

UDC 621. 3.06

On the role of key schedules in attacks on iterated ciphers / *V.I. Dolgov, A.A. Nastenko* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 7 – 15.

It is shown that the indices of DES cipher strength to attacks of the linear and differential cryptanalysis of the key deployment schemes are practically independent as it is for other indicators of well-known ciphers. In all schemes of keys unfolding the DES cipher after 16 cycles comes to the refractive random permutation corresponding degree. It is noted that the assertion that for any key diagrams and most of codes assigned to Markov, the distribution of differentials probability and linear hulls converges to the uniform distribution of the NIJ, after a certain number of cycles, is incorrect.

Tab. 4. Ref.: 18 items.

УДК 621. 3.06

S-блоки для современных шифров / *В. И. Долгов, Е.Д. Мельничук* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 16 – 23.

Приводятся результаты оценки эффективности S-блоковых конструкций, применяемых в современных блочных симметричных шифрах. Используется новый показатель оценки эффективности S-блоковых конструкций, который строится на основе оценки числа циклов шифра, необходимых ему для прихода к стационарному состоянию, свойственному случайной подстановке. Показывается, что с высокой вероятностью случайно выбранные S-блоки по этому показателю эффективности будут обладать криптографическими свойствами, не уступающими лучшим известным конструкциям. Предлагаются конструкции случайных S-блоков для шифра Rijndael и шифров, представленных на прошедший украинский конкурс по выбору национального стандарта шифрования

Табл. 13. Библиогр.: 7 назв.

УДК 621. 3.06

S-блоки для сучасних шифрів / *В.І. Долгов, Е.Д. Мельничук* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 16 – 23.

Наводяться результати оцінки ефективності S-блокових конструкцій, що застосовуються в сучасних шифрах. Використовується новий показник оцінки ефективності S-блокових конструкцій, який будується на основі оцінки числа циклів шифру, необхідних йому для приходу до стаціонарного стану, властивому випадкової підстановці. Показується, що з високою ймовірністю випадково вибрані S-блоки за цим показником ефективності будуть володіти криптографічними властивостями, що не

поступаються кращим відомим конструкціям S-блоків. Пропонуються конструкції випадкових S-блоків для шифру Rijndael і шифрів, представлених на минулий український конкурс з вибору національного стандарту шифрування.

Таб. 13. Бібліогр. : 7 назв.

UDC 621. 3.06

S-blocks for modern ciphers / *V.I. Dolgov, E.D. Melnichuk* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 16 – 23.

The results of evaluating the effectiveness of S-block structures used in modern ciphers variables. Use a new indicator assessing the effectiveness of S-block constructions, which is based on estimating the number of cycles of the cipher, it needs to come to steady state, characteristic of a random permutation. It is shown that with high probability of randomly selected S-boxes on the effectiveness of this indicator will have cryptographic properties, to the best-known designs S-boxes. Proposed construction of random S-boxes for the cipher Rijndael cipher, and, pre-represented in the previous Ukrainian national competition to select an encryption standard.

Tab. 13. Ref.: 7 items.

УДК 621.3.06

Криптоанализ шифра Mickey на основе анализа внутренних состояний / *А.В. Казимиров, Р.В. Олейников* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 24 – 30.

Рассмотрены генераторы гаммы с двумя внутренними регистрами, обновляемыми линейно и нелинейно, причем при тактировании состояния регистров являются взаимозависимыми (реализуется в виде взаимного управления). Рассматривается модель атаки, основанная на обратном тактировании генератора и анализа полученного дерева обратных состояний. Этот подход применяется к поточному шифру Mickey, полученные результаты сравниваются с известными теоретическими.

Табл. 7 Ил. 3. Библиогр.: 10 назв.

УДК 621.3.06

Криптоаналіз шифру Mickey на основі аналізу внутрішніх станів / *О.В. Казимиров, Р.В. Олійников* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 24 – 30.

Розглянуто генератори гами з двома внутрішніми регістрами, які оновлюються лінійно і нелінійно, причому при тактуванні стани регістрів є взаємозалежними (реалізується у вигляді взаємного управління). Розглядається модель атаки, заснована на зворотному тактуванні генератора і аналізі отриманого дерева зворотних станів. Цей підхід застосовується до потокового шифру Mickey, отримані результати порівнюються з відомими теоретичними.

Табл. 7. Іл. 3. Бібліогр.: 10 назв.

UDC 621.3.06

Mickey cipher cryptanalysis based on internal states analysis / *O.V. Kazymyrov, R.V. Oliynykov* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 24 – 30.

The key-stream generator with the internal state split into two parts, which are updated respectively linearly and nonlinearly is considered. Moreover, the state update function also depends on the internal state of the partner register implementing the so called self-mutual control. The attack scenario based on the reverse clocking of the generator and analysis of the obtained state backward tree is proposed. This approach with Mickey stream cipher is illustrated in practice. Own practical results are compared with the previous theoretical ones.

7 tab. 3 fig. Ref.: 10 items

УДК 621. 3.06

Вырожденные подстановки / *И.В. Лисицкая* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 31 – 38.

Исследуются дифференциальные и линейные свойства подстановок, названных вырожденными, под которыми понимаются подстановки, использование которых либо не позволяет в пределах ограниченного числа циклов, однозначно определенного для каждого шифра, достичь показателей случайной подстановки, либо стационарное значение, к которому приходит шифр, не соответствует ожидаемому, свойственному случайной подстановке. Приводится оценка вероятности случайного выбора подстановки вырожденного типа. Делается вывод, что попасть на вырожденную подстановку при случайном выборе является маловероятным событием. Этим подтверждается положение, являющееся основой предложенной новой методологии оценки стойкости блочных симметричных шифров к атакам дифференциального и линейного криптоанализа, состоящее в том, что все шифры после не-

большого начального числа циклов шифрования независимо от используемых S-блоков приобретают свойства случайных подстановок соответствующих степеней.

Табл. 7. Библиогр.: 12 назв.

УДК 621. 3.06

Вироджені підстановки / *І.В. Лисицька* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 31 – 38.

Досліджуються диференціальні й лінійні властивості підстановки названих виродженими, під якими розуміються підстановки, використання яких або не дозволяє в межах обмеженого числа циклів, однозначно визначеного для кожного шифру, досягти показників випадкової підстановки, або стаціонарне значення, до якого приходить шифр, не відповідає очікуваному, властивому випадкової підстановці. Наводиться оцінка ймовірності випадкового вибору підстановки виродженого типу. Робиться висновок, що потрапити на вироджения підстановку при випадковому виборі є дуже малоймовірним подією. Цим підтверджується положення, що є основою запропонованої нової методології оцінки стійкості блокових симетричних шифрів до атак диференціального та лінійного криптоаналізу, яке у тому, що всі шифри після невеликого початкового числа циклів шифрування незалежно від використовуваних S-блоків набувають властивостей випадкових підстановок відповідних степенів.

Таб. 7. Бібліогр. :12 назв.

UDC 621.3.06

Degenerated substitutions / *I.V. Lisitskaya* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 31 – 38.

The differential and linear properties of these degenerate substitutions, which are defined as the substitution are studied, their use either do not allow to achieve a random permutation in a limited number of cycles, which is uniquely defined for each cipher parameters, or steady-state value, which comes to code does not match the expected inherent a random permutation. The probability of a random choice of the substitution of degenerated type is estimated. It is concluded that to get to a degenerate substitution with a random choice is a very unlikely event. This confirms the position which is the basis for the proposed new methodology for assessing the strength of block symmetric ciphers to attacks of differential and linear cryptanalysis, which consists in the fact that all ciphers after a small initial number of cycles of encryption regardless of the S-boxes acquire the properties of random permutations of the powers.

Tab.7.Ref.:12 items.

УДК 681.3.06: 519.248.681

Про доказательство отсутствия эффективных байтовых дифференциальных характеристик для Rijndael-подобных шифров / *В.И. Руженцев* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 39 – 42.

Внимание сосредотачивается на уточнении представленных в одной из предыдущих работ леммы и теоремы об эффективных байтовых дифференциальных характеристиках, которые могут быть использованы в атаке усеченных дифференциалов. Представлены более корректная формулировка леммы о граничном числе активных колонок и новое доказательство теоремы об отсутствии эффективных байтовых дифференциальных характеристиках в Rijndael-подобных шифрах. Выполняется сравнение теоретически полученных оценок с полученными на практике результатами.

Табл. 1. Библиогр.: 10 назв.

УДК 681.3.06: 519.248.681

Про доведення відсутності ефективних байтових диференційних характеристик для Rijndael-подібних шифрів / *В.І. Руженцев* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 39 – 42.

Увага приділяється уточненню представлених в одній з попередніх робіт лемми та теореми стосовно ефективних байтових диференційних характеристик, які можуть бути використані в атаці усічених диференціалів. Представлено кореговане формулювання леми про граничну кількість активних колонок та нове доведення теореми про відсутність ефективних байтових диференційних характеристик в Rijndael-подібних шифрах. Виконується порівняння теоретично отриманих оцінок з отриманими на практиці результатами.

Табл. 1. Бібліогр.: 10 назв.

UDC 681.3.06: 519.248.681

About the proof of effective byte differential characteristics absence for Rijndael-like ciphers / *V.I. Ruzhentsev* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 39 – 42.

The attention is paid to the correction of the lemma and theorem presented in the previous paper. These statements are about the effective byte differential characteristics which can be used in truncated differential

attack. The correction of formula of lemma about the boundary number of active columns and the new proof of theorem about effective byte differential characteristics absence for Rijndael-like ciphers is presented. The comparison of theoretical results and practical estimations is carried out.

1 tab. Ref.: 10 items.

УДК 681.3.06

Формирование нелинейных узлов замен с использованием недвоичных криптографических функций / *А.А. Кузнецов, А.В. Коваленко, С.А. Исаев* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 43 – 57.

Исследуются математические модели регулярных нелинейных узлов замен (S-блоков) симметричных криптографических алгоритмов. Рассматривается традиционный подход, в котором для описания внутренней структуры S-блоков используется совокупность компонентных булевых функций. Исследуется аналитический аппарат недвоичных криптографических функций, с использованием которого разрабатывается математическая модель регулярных нелинейных узлов замен для формализации описания внутренней структуры S-блоков, развития методов их синтеза и оценки криптографических свойств. Приводятся результаты вычислительного поиска S-блоков с использованием предлагаемого математического аппарата, показано, что по нелинейности и автокорреляции сформированные узлы замен обладают улучшенными свойствами.

Табл. 4. Библиогр.: 14 назв.

УДК 681.3.06

Формування нелінійних вузлів замін з використанням недвійкових криптографічних функцій / *О.О. Кузнецов, О.В. Коваленко, С.О. Ісаєв* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 43 – 57.

Досліджуються математичні моделі регулярних нелінійних вузлів замін (S-блоків) симетричних криптоалгоритмів. Розглядається традиційний підхід, у якому для опису внутрішньої структури S-блоків використовується сукупність компонентних мулевих функцій. Досліджується аналітичний апарат недвійкових криптографічних функцій, з використанням якого розробляється математична модель регулярних нелінійних вузлів замін для формалізації опису внутрішньої структури S-блоків, розвитку методів їх синтезу та оцінки криптографічних властивостей. Доводяться результати обчислювального пошуку S-блоків з використанням пропонує мого математичного апарату, показано, що по нелінійності та автокореляції сформовані вузли замін володіють покращеними властивостями.

Табл. 4. Бібліогр.: 14 назв.

UDC 681.3.06

Synthesis of nonlinear substitution components with the use of non-binary cryptographic functions // *A.A. Kuznetsov, A.V. Kovalenko, S.A. Isaev* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 43 – 57.

The mathematical model of regular nonlinear substitution components (S-blocks) of symmetric cryptographic algorithms is studied. The traditional approachis considered, in which the set of component Boolean functionsis used to describethe S-boxesinternal structure. The analytical apparatus of non-binary cryptographic functions with the use of which the mathematical model of regular nonlinear substitution component is developed to formalize the description of the internal structure of S-blocks, development of methods for their synthesis and evaluation of their cryptographic properties are investigated.The results of computational searchof S-blocks using the mathematical apparatus are shown, it is demonstrated that the nonlinearity and autocorrelation of the generated S-blocks have improved properties.

Tab.4. Ref.: 14 items.

КРИПТОГРАФИЧЕСКИЕ ПРЕОБРАЗОВАНИЯ
CRYPTOGRAPHIC TRANSFORMATION

УДК 004.056.55

Исследование методов вычисления инверсии в алгоритме NTRU / *Е.Г. Качко, Д.С. Балагура, К.А. Погребняк, Ю.И. Горбенко* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 58 – 63.

Цель работы – исследование различных методов вычисления инверсии, их оптимизация с учетом возможностей параллельных вычислений. Исследуются алгоритмы инверсии, приведенные в различных источниках как для инверсии в общем случае, так и для инверсии для модуля 2. Рассмат-

ривается возможность оптимизации указанных алгоритмов с точки зрения их возможной параллелизации. Показывается возможность существенного ускорения алгоритмов инверсии за счет их параллелизации и других методов ускорения алгоритмов.

Табл. 3. Ил. 1. Библиогр.: 5 назв.

УДК 004.056.55

Дослідження методів обчислення інверсії у алгоритмі NTRU / *О.Г. Качко, Д.С. Балагура, К.А. Погребняк, Ю.І. Горбенко* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 58 – 63.

Мета роботи – дослідження різноманітних алгоритмів обчислення інверсії, їх оптимізація з урахуванням можливостей паралельних обчислень. Досліджуються алгоритми інверсії, наведені у різних джерелах як для інверсії у загальному випадку, так і для інверсії для модуля 2. Розглядається можливість оптимізації зазначених алгоритмів з точки зору їх можливої паралелізації. Показується можливість суттєвого прискорення алгоритмів інверсії за рахунок їх паралелізації та інших методів прискоренні алгоритмів

Таб. 3 Іл. 1. Бібліогр. :5 найм.

UDC 004.056.55

Investigation of methods for calculation of the inversion algorithm, NTRU / *E.G. Kachko, D.S. Balagura, K.A. Pogrebnyak, Y.I. Gorbenko* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 58 – 63.

The purpose of this article is to investigate different methods for calculating the inversion, to optimize them with the possibilities of parallel computing. The inversion algorithms presented in various sources for the inversion in the general case, and for the inversion for module 2 are studied. The possibility of the indicated algorithms optimization from the point of view of their possible parallelization is considered. The possibility of significant acceleration of inversion algorithms due to their parallelization, and other methods to accelerate algorithms is shown.

Tab 3. Fig.1. Ref.: 5 items.

УДК 681. 3.06

Криптографическая поддержка услуг защиты от несанкционированного доступа / *Ю.М. Леншина, А.В. Леншин* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 64 – 71.

Проанализированы требования национальных и международных документов, регламентирующих защиту от несанкционированного доступа. Разработаны рекомендации по усовершенствованию НД ТЗИ в части спецификации услуг защиты от угроз приватности. Рассмотрены подходы к криптографической поддержке услуг защиты от несанкционированного доступа и методология оценки уровня стойкости услуг безопасности. Приведены условия обеспечения доказуемого уровня стойкости.

Табл. 5. Ил. 1. Библиогр.: 9 назв.

УДК 681. 3.06

Криптографічна підтримка послуг захисту від несанкціонованого доступу / *Ю.М. Лєншина, А.В. Лєншин* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 64 – 71.

Проаналізовано вимоги національних та міжнародних документів, що регламентують захист від несанкціонованого доступу. Розроблено рекомендації з вдосконалення НД ТЗІ в частині специфікації послуг захисту від загроз порушення приватності. Розглянуті підходи до криптографічної підтримки послуг захисту від несанкціонованого доступу та методологія оцінки рівня стійкості послуг безпеки. Наведено умови забезпечення доказового рівня стійкості.

Таб. 5. Іл. 1. Бібліогр. : 9 назв.

UDC 681. 3.06

Cryptographic support of unauthorized access protection services / *I.M. Lyenshyna, A.V. Lyenshyn* // Radiotekhnika: All-Ukr. Sci. Interdep. Mag. 2012. N 171. P. 64 – 71.

The requirements of national and international instruments governing the protection from unauthorized access are analyzed. Recommendations for improving ND TZI services regarding the privacy protection are proposed. The approaches to the cryptographic support of unauthorized access protection services and evaluation methodology of its strength level are discovered. Conditions to provable strength of cryptographic support of privacy services are given

Tab.5. Fig. 1. Ref.: 9 items.

УДК 681.3.06

Модель использования облачных вычислений для задач асимметричного криптоанализа на примере факторизация чисел методом ρ-Полларда / *К.А. Погребняк, Д.В. Повтарев* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 72 – 78.

Приводится анализ преимуществ и недостатков использования облачных вычислений для задач криптонализа, на основе которых формируется перечень требований к криптоаналитической системе. Разрабатывается общая модель использования облачных вычислений для задач асимметричного криптоанализа.

Приводится описание модуля, использующего облачные вычисления, и проводится анализ результатов его использования в сравнении с аналогичным решением, функционирующим на базе персонального компьютера.

Табл. 2. Ил. 2. Библиогр.: 7 назв.

УДК 681.3.06

Модель використання хмарних обчислень для задач асиметричного криптоаналізу на прикладі факторизації чисел методом ρ-Полларда / *К.А. Погребняк, Д.В. Повтарев* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 72 – 78.

Наводено аналіз переваг та недоліків використання хмарних обчислень у задачах криптоналізу, на основі яких формується перелік вимог до криптоаналітичної системи. Описується розроблена загальна модель використання хмарних обчислень для задач асиметричного криптоаналізу.

Наводено опис розробленого модуля, що використовує ресурси на основі хмарних обчислень та проводиться аналіз результатів його використання у порівнянні з аналогічним рішенням, що функціонує на базі персонального комп'ютера.

Табл. 2. Іл. 2. Бібліогр.: 7 назв.

UDC 681.3.06

Model of using cloud computing for problems of asymmetric cryptanalysis illustrated by the example of factorization of numbers by ρ-Pollard algorithm / *K.A. Pogrebnyak, D.V. Povtariev* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 72 – 78.

The main advantages and disadvantages of using the cloud computing for the problems of cryptanalysis were analyzed. The list of requirements for cryptanalytic system and general model based on it was designed according to the environment specific.

The program module of the cloud computing system for cryptanalysis tasks was developed and results of its usage were analyzed and compared with the same solution functioning on the PC basis.

Tab. 2. Fig. 2. Ref.: 7 items.

УДК 004.056.55

Анализ стойкости вычислительных задач, основанных на билинейных отображениях / *И.Д. Горбенко, Л.В. Макутонина* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 79 – 89.

Одним из самых перспективных направлений развития криптосистем на идентификаторах является объединение математики спаривания точек на эллиптических кривых с алгебраическими решетками. Представлены определения и анализ основных вычислительных задач, основанных на билинейных отображениях, в последующих работах будет представлен анализ вычислительных задач на решетках, а также объединенной математики.

Библиогр: 17 назв.

УДК 004.056.55

Аналіз стійкості обчислювальних задач, що засновані на білінійних відображеннях / *І.Д. Горбенко, Л.В. Макутоніна* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 79 – 89.

Одним із найперспективніших напрямів розвитку криптосистем на ідентифікаторах є об'єднання математики спарювання точок на еліптичній кривій з алгебраїчними решітками. У статті надано визначення та аналіз основних обчислювальних задач, що засновані на білінійних відображеннях, у наступних роботах буде надано аналіз обчислювальних задач на решітках, а також об'єднаної математики.

Бібліогр: 17 назв.

UDC 004.056.55

Resistance analysis computational problems based on the bilinear maps / *I.D. Gorbenko, L.V.Makutonina* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 79 – 89.

At the moment, one of the most promising areas of identity-based cryptosystems is the union of mathematics pairings points on elliptic curves and algebraic lattices. This paper presents the definition and analysis of the major computing tasks based on bilinear maps. The subsequent works will analyze the computational problems on lattices and combined math.

Ref.: 17 items.

УДК 621. 391:519.2:519.7

Методы противодействия атакам специального вида на схемы направленного шифрования в кольцах срезанных полиномов / *Д.В. Иваненко* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 90 – 98.

Анализируется методы противодействия атакам на реализацию, которые основываются на анализе энергопотребления. Рассматривается реализация схем направленного шифрования в кольце срезанных полиномов, принимая во внимания встраивания методов противодействия, таких как рандомизация и «заслепление». Были выявлены недостатки и преимущества методов противодействия относительно таких атак, как SPACPAи DPA, которые были направлены на схему направленного шифрования.

Табл.1 Ил. 5 Библ.: 19 назв.

УДК 621.391:519.2:519.7

Методи протидії атакам спеціального виду на схеми направленого шифрування у кільцях зрізаних поліномів / *Д.В. Іваненко* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 90 – 98.

Аналізуються існуючі методи протидії атакам на реалізацію, які базуються на аналізі енергоспоживання. Розглядається реалізація схеми направленого шифрування, беручи до уваги впровадження методів протидії, таких яке рандомізація та засліплення. Було виявлено недоліки та переваги методів протидії відносно таких атак спеціального виду як SPA,CPA та DPA, які були направлені на схему направленого шифрування.

Табл. 1. Іл. 5. Бібл.: 19 назв.

UDC 621. 391:519.2"519.7

Methods to counteract side channel attacksto the schemes in rings of truncated polynomials / *D.V.Ivanenko* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 90 – 98.

The existing methods to counteract the attacks on the implementations based on the energy consumption analysis are considered. Realization of the directed encryption schemes in rings of truncated polynomials, given the embedding methods to counteract such as randomization and "blindness", is analyzed. The advantages and disadvantages of methods to counteract SPA CPA and DPA attacks sent to the directed encryption schemes are identified.

Tabl. Fig.5. Ref.: 19items.

УДК 681.3.06

Сущность и анализ криптографических требований стандарта NISTSP 800-90B / *И.Д. Горбенко, Р.И. Мордвинов* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 99 – 108.

Одним из наиболее важных моментов в криптографии являются ключевые данные. Они должны удовлетворять ряду выдвигаемых к ним требований, одним из которых является случайность последовательности. Для удовлетворения данным свойствам, ключевые данные принято создавать с помощью генераторов случайной последовательности. В статье рассматривается стандарт, описывающий требования и методы тестирования таких генераторов и их компонентов.

Ил. 1. Библиогр: 1 назв.

УДК 681.3.06

Сутність та аналіз криптографічних вимог стандарту NISTSP 800-90B / *Ю.І. Горбенко, Р.І. Мордвінов* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 99 – 108.

Одним з найбільш важливих моментів у криптографії є ключові дані. Вони повинні задовольняти ряду вимог, що до них висуваються, одним з яких є випадковість послідовності. Для задовільнення даним вимогам, ключові дані прийнято створювати за допомогою генераторів випадкової послідов-

ності. В статті розглядається стандарт, в якому описані вимоги та методи тестування таких генераторів та їх компонентів.

Іл. 1. Бібліогр: 1 назв.

UDC 681.3.06

Essence and analysis of cryptographic requirements from NIST SP 800-90B / *U.I. Gorbenko, R.I. Mordvinov* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 99 – 108.

Thekey dataare one of the mostimportant issuesin cryptography.They have tomeet a numberof requirementsput forwardto them, one of which is a randomsequence.To satisfythese properties it is a coomon practice rogeneratekey datausinga random sequencegenerator. Thestandard describingthe requirementsand test methods forsuchgenerators andtheir componentsn are considered.

1 fig. Ref.: 1 items.

УДК 681.3.06

Универсальное хеширование с ограничением функционального поля алгебраических кривых / *Е.В. Котух* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 109 – 115.

Представлено решение задачи построения универсального хеширования с ограничением функционального поля алгебраических кривых. Рассмотрены свойства хеширования по двум параметрическим базисам кривой Судзуки. Получены оценки вероятности коллизии, сложности вычислений и ключевого пространства для хеширования. Универсальное хеширование с ограничением функционального поля алгебраической кривой приводит к существенному снижению сложности вычислений. При этом уменьшается число хешируемых данных. Требуется оптимизация базиса функционального поля. Эффект достигается на сложных многопараметрических кривых.

Табл. 1. Библиогр.: 9 назв.

УДК 681.3.06

Універсальне гешування з обмеженням функціонального поля алгебраїчних кривих / *Є.В. Котух* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 109 – 115.

Представлено рішення задачі побудови універсального гешування з обмеженням функціонального поля алгебраїчних кривих. Розглянуто властивості гешування по двох параметричних базисах кривої Судзукі. Отримано оцінки ймовірності колізії, складності обчислень і ключового простору для хешування. Універсальне гешування з обмеженням функціонального поля алгебраїчної кривої призводить до істотного зниження складності обчислень. При цьому зменшується число гешированих даних. Потрібно оптимізація базису функціонального поля. Ефект досягається на складних багато параметричних кривих.

Табл. 1. Бібліогр.: 9 назв.

UDC 681.3.06

Universal hashing with limited functional field of algebraic curves / *Ye. Kotukh* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 109 – 115.

The solution of the problem of constructing a universal hashing with limited functional field of algebraic curves is presented. The properties of the hashing on two-parametric basis of Suzuki curve are considered. Estimation of the collision probability, computational complexity and key space for hashing are obtained. Universal hashing with the limited function field of the algebraic curve results in a significant reduction in the computational complexity. It causes that the number of hash data is decreased. Optimization of the functional field basis is needed. Such effect can be achieved on complex multi-parameter curves.

Tab.1. Refs.: 9 titles.

УДК 004.056.55

Алгоритм решета числового поля / *И. Д. Горбенко, М.В. Есина* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 116 – 122.

Определены особенности факторизации на основе общего алгоритма «решета числового поля», а также обоснован выбор факторных баз. Определён принцип построения алгебраической факторной базы и алгебраической базы характеров. Также определена оценка сложности решета числового поля.

Библиогр.: 5 назв.

УДК 004.056.55

Алгоритм решета числового поля / *І. Д. Горбенко, М. В. Єсіна* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 116 – 122.

Визначено особливості факторизації на основі загального «решета числового поля», а також обгрунтовано вибір факторних баз. Визначено принцип побудови алгебраїчної факторної бази та ал-

гебраїчної бази характерів. Також визначена оцінка складності решета числового поля.

Бібліогр.: 5 назв.

UDC 004.056.55

Number field sieve algorithm / *I. D. Gorbenko, M. V. Yesina* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 116 – 122.

The major features of the factorization based on a total "number field sieve" are defined, and the choice of factor bases is justified. The principle of algebraic factor base and algebraic characters base are defined. The estimate of the number field sieve is determined as well.

Ref.: 5 items

УДК 004.056.55

Подходы к распараллеливанию программной реализации операции умножения целых чисел / *В.Ю.Ковтун, А.А.Охрименко* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 123 – 131.

Рассматриваются подходы к повышению производительности программной реализации алгоритма умножения целых чисел для 32- и 64-разрядных платформ, посредством технологии распараллеливания. Основная идея распараллеливания алгоритма, состоит в использовании механизма отложенного переноса предложенного авторами ранее. Отложенный перенос позволяет избавиться от связности в итерациях циклов накопления суммы произведений, что позволяет выполнить их параллельно. После завершения потоков накопления суммы произведений, производится корректировка конечного результата посредством учета переносов. Такие подходы увеличивают общее число операций сложения, однако уменьшают общее время выполнения программной реализации на многопроцессорных и многоядерных вычислительных системах.

Табл. 1. Ил. 4. Библиогр.: 11 назв.

УДК 004.056.55

Підходи до розпаралелювання програмної реалізації операції множення цілих чисел / *В.Ю. Ковтун, А.О. Охріменко* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 123 – 131.

Розглядаються підходи до підвищення продуктивності програмної реалізації алгоритму множення цілих чисел для 32- і 64-розрядних платформ, за допомогою технології розпаралелювання. Основна ідея розпаралелювання алгоритму полягає у використанні механізму відкладеного перенесення запропонованого авторами раніше. Відкладене перенесення дозволяє позбавитися від зв'язності в ітераціях циклів накопичення суми добутків, що дозволяє виконати їх паралельно. Після завершення потоків накопичення суми добутків, проводиться корегування кінцевого результату за допомогою врахування переносів. Такі підходи збільшують загальне число операцій додавання, однак зменшують загальний час виконання програмної реалізації на багатопроцесорних і багатоядерних обчислювальних системах.

Табл. 1. Іл. 4. Бібліогр.: 11 назв.

UDC 004.056.55

Approaches to parallelization of software implementation of integer multiplication / *V.Y. Kovtun, A.O. Okhrimenko* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 123 – 131.

Several approaches to the increasing performance of software implementation of integer multiplication algorithm for the 32-bit & 64-bit platforms via parallelization are considered. The main idea of algorithm parallelization consists in delayed carry mechanism using proposed by the authors earlier [11]. The delayed carry allows to get rid of connectivity in loop iterations for sums accumulation of products, which allows parallel execution of loops iterations in separate threads. After the finishing sum accumulation threads, it is necessary to make corrections in final result via carries considerations. First approach consists in parallelization optimization for the two execution threads and second approach is an evolution of the first approach and oriented on at most three execution threads. Proposed approaches for parallelization allows increasing the total algorithm computational complexity, as for one execution thread, but decrease total execution time on multi-core CPU.

1 tab. 4 fig. Ref.: 11 items.

УДК 004.056.5

Аутентификция аппаратных средств КЗИ / *А.А. Шевчук, Ю.И. Горбенко* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 132 – 139.

Рассматриваются уточнения в модели угроз средств КЗИ. Рассматривается влияние особенности использования ключей сессии в ЭЦП, основанные на схеме Ниберг – Рюпеля и DSA. Рассматривается

возможное влияние третьей стороны, участвующей в распространении средства. Предлагается обязательная аутентификация средств КЗИ, которые предоставляют услугу ЭЦП по определенным алгоритмам. Предлагается протокол аутентификации, с привлечением производителя средства.

Табл. 1. Библиогр.: 8 назв.

УДК 004.056.5

Автентифікація апаратних засобів КЗІ / *О.А. Шевчук, Ю.І. Горбенко* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 132 – 139.

Розглядаються уточнення до моделі загроз засобів КЗІ. Розглядається вплив особливості використання ключів сесії в ЕЦП, що грунтуються на схемі Ніберг – Рюпеля та DSA. Розглядається можливий вплив третьої сторони, що бере учать у розповсюдженні засобу. Пропонується обов'язкова автентифікація засобів КЗІ, що надають послугу ЕЦП за визначеними алгоритмами. Пропонується протокол автентифікації, с залученням виробника засобу.

Табл. 1. Библиогр.: 8 назв.

UDC 004.056.5

HSM Authentication / *O.A.Shevchuk, U.I. Gorbenko* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 132 – 139.

Clarification of the HSM threat modelsis considered. The influence of the particular use of session keys in the digital signature schemes based on Nyberg-Ryupel and DSA algorithms can be used by an intruder. HSM authentication protocol is proposed to reduce threats.

Tab. 1. Refs.: 8 titles.

УДК 681.3.06

Универсальное хеширование по обобщенным кривым Гурвица / *О.Г. Халимов, А.Н. Герцог* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 140 – 146.

Представлено решение задачи построения кривых Гурвица с наибольшим отношением числа точек к роду на основе метода приведения к обобщенным кривым минимального рода с перебором значений наибольшего показателя степени кривой. Рассмотрены свойства метода приведения к обобщенным кривым Гурвица минимального рода, примеры построения кривых по делителям порядка поля. Показано, что предложенный метод расширяет множество кривых, так как определяет построение кривых по одному делителю порядка поля и в среднем в 7.46 раза быстрее по сравнению с перебором по наименьшему показателю степени кривой.

Библиогр.: 10 назв.

УДК 681.3.06

Універсальне гешування за узагальненими кривими Гурвіца / *О.Г. Халімов, О.М. Герцог* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 140 – 146.

Представлено рішення задачі побудови кривих Гурвіца з найбільшим відношенням числа точок до роду на основі методу приведення до узагальнених кривим мінімального роду з перебором значень найбільшого показника ступеня кривої. Розглянуто властивості методу приведення до узагальнених кривим Гурвіца мінімального роду, приклади побудови кривих по дільникам порядку поля. Показано, що запропонований метод розширює безліч кривих, оскільки визначає побудову кривих по одному дільнику порядку поля і в середньому в 7.46 рази швидше у порівнянні з перебором за найменшим показником ступеня кривої.

Бібліогр.: 10 назв.

UDC 681.3.06

Universal hashing by generalized Hurwitz curves / *O.G. Khalimov, A.N. Guetzog* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 140 – 146.

The solution of the problem of Hurwitz curves construction with the highest ratio of the number of points to the genus on the basis of the method of reduction to the generalized curves of minimum genus with listing of the curve greatest exponent values is presented. The properties of the method of reduction to the generalized Hurwitz curves of minimum genus, some examples of curves construction by divisor of the field order are considered. It is shown that the proposed method increases the number of curves, since it determines the construction of curves, one factor of the order of the field and on the average is 7.46 times faster as compared to the listing of the smallest exponent of the curve.

Ref.: 10 items.

УДК 621.391:519.2:519.7

Оценка стойкости направленного шифра NTRU к атаке с адаптивно подобранными шиф-

ротекстами / *А.П. Бубырь, И.Д. Горбенко* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 147 – 151.

Представлены результаты исследования программной модели атаки с адаптивно подобранными шифротекстами на алгоритм направленного шифрования NTRU. Экспериментально определена зависимость вероятности ошибочного дешифрования сообщения от использванных общесистемных параметров. Установлено, что для гарантированного восстановления открытого текста без знания личного ключа достаточно иметь не более 30-40 криптограмм специального вида. Даны рекомендации по защите от подобных атак. Сделаны выводы о допустимых сферах применения данного алгоритма.

Ил. 3. Библиогр.: 3 назв.

УДК 621.391:519.2:519.7

Оцінка стійкості направленого шифру NTRUдо атаки з адаптивно підібраними шифротекстами / *А.П. Бубир, І.Д. Горбенко* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 147 – 151.

Представлено результати дослідження програмної моделі атаки з адаптивно підібраними шифротекстами на алгоритм направленого шифрування NTRU. Експериментально визначена залежність ймовірності помилкового дешифрування повідомлення від використаних загальносистемних параметрів. Встановлено, що для гарантованого відновлення відкритого тексту без знання особистого ключа достатньо мати не більше, ніж 30-40 криптограм спеціального виду. Надано рекомендації щодо захисту від подібних атак. Зроблені висновки про допустимі сфери використання даного алгоритму.

Іл. 3. Бібліогр.: 3 назви.

UDC621.391:519.2:519.7

Evaluation of lattice-based public key algorithm NTRU resistance against adaptive chosen ciphertexts attack / *A.P. Bubyr, I.D. Gorbenko* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 147 – 151.

Results of research intothe adaptive chosen ciphertexts attack's program model are given. Dependence of incorrect message recovery probability on the used system-wide parameters was defined experimentally. It was found that for the guaranteed recovery of plaintext without knowledge of the private key it is enough to have at most 30-40 cryptograms ofa special kind. Recommendations for the protection against these attacks are provided. The conclusions about the allowable scope of the algorithm are considered.

1fig.,Ref.: 3 items.

УДК 621.391:519.2:519.7

Анализ двух методов аутентификации человека для применения в электронном биометрическом паспорте / *К.А. Бугаенко, И.Д. Горбенко* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 152 – 158.

Изучаются стандартные механизмы аутентификации личности по геометрии лица, по динамической подписи лица, их возможности, преимущества и недостатки. Проводится анализ методов аутентификации лица и предложения по их совершенствованию. Как и анализ по личной рукописной подписи, определение идентичности по фотографии в паспорте гражданина Украины относят к наиболее доступного и признанному методу распознавания личности. Применение двух методов аутентификации личности по биометрическим признакам геометрии лица и по рукописному (динамическом) подписи в паспорте гражданина Украины будут намного эффективнее, если применять их одновременно. Так как вероятность подделки биометрических данных уменьшается в несколько раз по сравнению с отдельным применением каждого из методов. Одновременно уменьшатся и ошибки как первого, так и второго рода, что приведет к повышению эффективности и надежности рассматриваемых двух методов.

Табл. 3. Ил.: 4. Библиогр.: 6 назв.

УДК 621.391:519.2:519.7

Аналіз двох методів автентифікації особи для застосування в електронному біометричному паспорті / *Х.А. Бугаєнко, І.Д. Горбенко* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 152 – 158.

Вивчаються стандартні механізми автентифікації особи по геометрії обличчя, по динамічному підпису особи, їх можливості, переваги та недоліки. Та проводиться аналіз методів автентифікації особи та пропозиції з їх вдосконалення. Як і аналіз по особистому рукописному підпису, визначення ідентичності по фотографії в паспорті громадянина України відносять до найбільш доступного і визнаному методу розпізнавання особистості. Застосування двох методів для автентифікації особи по біометричним ознакам геометрії обличчя та по рукописному (динамічному) підписі в паспорті громадянина України будуть набагато ефективніші, якщо застосовувати їх одночасно. Так як ймовірність підробки біометричних даних зменшується в декілька разів в порівняні з окремим застосуванням ко-

жного з методів. Одночасно зменшаться й помилки як першого, так і другого роду, що призведе до підвищення ефективності і надійності розглянутих двох методів.

Табл. 3. Іл.: 4. Бібліогр.: 6 назв.

UDC 621.391:519.2:519.7

Analysis of two methods of a person authentication for the use in electronic biometric passports / *K.A. Bugaenko, I.D. Gorbenko* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 152 – 158.

The standard authentication mechanisms of a person by the face geometry, dynamic signature of a person, their capabilities, advantages and disadvantages are considered. Analysis of the authentication methods and proposals for their improvement are given. Like the analysis by a personal handwritten signature, the identity definition by a photo in the passport of a citizen of Ukraine belongs to the most affordable and recognized methods of a person recognition. The use of two methods to authenticate a person by biometric characteristics and geometry of the face on the manuscript (dynamic) signature in the passport of a citizen of Ukraine will be much more effective if we use them simultaneously. The probability to falsify the biometric data is reduced several times as compared with the individual usage of each method. Simultaneously, the errors both of the first and the second kind are reduced, all these will result in higher efficiency and reliability of the considered two methods.

Tab.3. Fig.4. Ref.: 6 items

ПЕРЕДАЧА И ОБРАБОТКА ИНФОРМАЦИИ
INFORMATION TRANSMISSION AND PROCESSING

УДК 621.391:519.2:519.7

Анализ методов обезличивания персональных данных / *Ю.И. Горбенко, А.С. Тоцкий, В.А. Пономарь* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 159 – 163.

Проанализировано методы обезличивания данных, которые удовлетворяют требованиям защиты персональных данных. Приведен алгоритм одного из методов и предложены разные варианты его реализации с соответствующими приоритетами: быстродействие, используемая память, сохранение необходимых статистических свойств.

Табл. 3. Библиогр.: 5 назв.

УДК 621.391:519.2:519.7

Аналіз методів знеособлення персональних даних / *Ю.І. Горбенко, О.С. Тоцький, В.А. Пономар* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 159 – 163.

Проаналізовано методи знеособлення даних, які задовольняють вимогам захисту персональних даних. Наведено алгоритм одного з методів та запропоновані різні варіанти його реалізації з відповідними пріоритетами: швидкодія, необхідна пам'ять та збереження необхідних статистичних властивостей.

Табл. 3. Бібліогр.5 назв.

UDC 621.391:519.2:519.7

Analysis of the methods of personal data anonymization / *Y.I. Gorbenko, A.S. Totsky, V.A. Ponomar* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 159 – 163.

The data anonymization techniques, which satisfy the requirements of the protection of personal data, is analyzed. An algorithm of one of the methods is given and different options for its implementation are proposed with the relevant priorities: performance, memory usage, conservation of the necessary statistical properties.

3 tab. Ref.: 5 items.

УДК 681.3.06

Модель институционального управления деятельностью по защите информации / *А.В. Потий, Д.Ю. Пилипенко* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 164 – 170.

Предлагается модель институционального управления деятельностью по защите информации. Проведено онтологическое моделирование института информационной безопасности и предметной области культуры информационной безопасности. Рассмотрены ключевые субъекты деятельности по защите информации: центр безопасности и агент безопасности. Установлены связи между субъектами деятельности по защите информации и компонентами модели.

Ил. 2. Библиогр.: 4 назв.

УДК 681.3.06

Модель інституціонального управління діяльністю із захисту інформації / *О.В. Потій, Д.Ю. Пилипенко* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 164 – 170.

Запропоновано модель інституціонального управління діяльністю із захисту інформації. Проведено онтологічне моделювання інституту інформаційної безпеки та предметної області культури інформаційної безпеки. Розглянуто головні суб'єкти діяльності із захисту інформації: центр безпеки та агента безпеки. Встановлено зв'язок між суб'єктами діяльності із захисту інформації та компонентами моделі.

Іл. 2. Бібліогр.: 4 назви.

UDC 681.3.06

Institutional model of information security activities management / *A.V. Potiy, D.Y. Pilipenko* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 164 – 170.

An institutional model of information security activities management is proposed. An ontological modeling of information security institution and subject domain of information security culture is carried out. The main subjects of information security activities are considered, namely the security center and security agent. The interconnections between the subjects of the information security activities and other components of this model are established.

2 fig. Ref.: 4 items.

УДК 681.323

Количественная оценка уязвимостей информационно-телекоммуникационных систем / *А.А. Замула, С.А.Сирота, Н.И. Косиковская* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 171 – 176.

Проведен анализ архитектуры информационных систем, уязвимостей, которые присущи данным системам, и которые могут быть использованы злоумышленниками при проведении атак на такие системы. Получена математическая модель оценки уязвимостей информационно-телекоммуникационных систем.

Библиогр.: 3 назв.

УДК 681.323

Кількісна оцінка вразливостей інформаційно-телекомунікаційних систем / *О.А.Замула, С.О. Сирота, Н.І. Косіковська* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 171 – 176.

Проведений аналіз архітектури інформаційних систем, вразливостей, які можуть бути притаманні даним системам, і які можуть бути використані зловмисниками при реалізації атак на такі системи. Отримана математична модель оцінки вразливостей інформаційно-телекомунікаційних систем.

Бібліогр.:3 назв.

UDC 681.323

Quantitative assessment of information technology systems vulnerabilities / *O.A.Zamula, S.O.Syrota, N.I.Kosikovskaia* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 171 – 176.

The analysis of information systems architecture, vulnerabilities that are inherent in these systems, and that can be used by intruders to conduct attacks on such systems was carried out. The mathematical model of vulnerability assessment of information technology systems was obtained.

Ref.: 3 items.

УДК 621.391

Предложения по построению широкополосных систем передачи со сложными сигналами / *А.А. Замула* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 177 – 184.

Сформулирована задача выбора сложных сигналов для широкополосных систем передачи. Определены необходимые и достаточные условия построения абсолютно стойкой системы передачи на уровне источника сложных сигналов

Ил. 3. Библиогр.: 4 назв.

УДК 621.391

Пропозиції щодо побудови широкополосних систем передачі зі складними сигналами / *О.А. Замула* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 177 – 184.

Сформульовано задачу відбору складних сигналів для широкополосних систем передачі інформації. Визначені необхідні та достатні умови побудови абсолютно стійких систем передачі на рівні джерела складних сигналів.

Іл. 3. Бібліогр.: 4 назв.

UDC 621.391

Proposal for construction of the wide-band transmission systems with complex signals / *O.A. Zamula* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 177 – 184.

The task of choosing complex signals for wideband transmission systems is formulated. The necessary and sufficient conditions for constructing transmission systems that are completely resistant on the level of complex signals' source are defined.

3 fig. Ref.: 4 items.

УДК 004.056.5: 004.775

Синтез системы многоуровневой защиты корпоративных порталов на платформе MS SHARE POINT / *Б.О. Бабич, А.В. Сагун, О.А. Кожуховская, А.Д. Кожуховский* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 185 – 188.

Рассмотрены вопросы безопасности корпоративных порталов, современные тенденции, предложено практическое решение системы защищённого корпоративного портала на базе многоуровневой системы защиты корпоративных порталов платформы MS Share Point при помощи интегрируемого программного обеспечения Microsoft. Представлена универсальная тестовая топология для изучения алгоритма построения защищенной облачной системы.

Ил.2 . Библиогр.: 7 назв.

УДК 004.056.5: 004.775

Синтез системи багаторівневого захисту корпоративних порталів на платформі MS SHARE POINT / *Б.О. Бабич, А.В. Сагун, О.А. Кожухівська, А.Д. Кожухівський* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 185 – 188.

Розглянуто питання безпеки корпоративних порталів, сучасні тенденції і одне із практичних рішень. Досліджено питання побудови багаторівневої системи захисту корпоративних порталів на платформі MS SHARE POINT за допомогою інтегруємого програмного забезпечення Microsoft. Представлена універсальна тестова топологія для вивчення алгоритму побудови захищеної хмарової системи.

Ключові слова: корпоративний портал, хмарові системи, безпека, багаторівневий захист, Share Point.

Іл. 2. Бібліогр.: 7 назв.

UDC 004.056.5: 004.775

Synthesis of system of multilevel protection of corporate portals on the base of Share Point Platform / *B.A. Babich, A.V. Sagun, O.A. Kozhukhovska, A.D. Kozhukhovskiy* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 185 – 188.

The problems of corporate portals' safety, modern trends and a practical solution were considered. The problems of building a multilevel system of corporate portals' protection on the base of MS Share Point Platform with the help of integrated Microsoft Software were investigated. The universal test topology for studying the algorithm of building up the protected cloudy system was presented.

Fig.2. Ref.: 7 items.

УДК 621.396

Критерии и показатели эффективности стеганографических систем защиты информации / *А.А. Смирнов* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 189 – 197.

Рассматривается задача стеганографической защиты информации, основной целью которой является сокрытие не только смыслового содержания передаваемых данных, но и самого факта осуществления скрытной передачи. Исследуется математическая модель и структурная схема стеганографической системы. Вводятся основные операторы, формально описывающие основные вычислительные процессы при стенографическом преобразовании информации на передающей и принимающей стороне. Вводятся критерии и показатели эффективности функционирования стеганографической системы. В общем виде формализуется задача совершенствования стеганосистемы через максимизацию обобщенного показателя ее эффективности.

Ил. 3. Библиогр.: 10 назв.

УДК 621.396

Критерії й показники ефективності стеганографічних систем захисту інформації / *А.А. Смірнов* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 189 – 197.

Розглядається завдання стеганографічного захисту інформації, основною метою якого є приховання не тільки визначеного змісту переданих даних, але й самого факту здійснення потайливої пере-

дачі. Досліджується математична модель і структурна схема стеганографічної системи. Уводяться основні оператори, що формально описують основні обчислювальні процеси при стенографічному перетворенні інформації на передавальній і приймаючій стороні. Уводяться критерії й показники ефективності функціонування стеганографічної системи. У загальному виді формалізується завдання вдосконалювання стеганосистеми через максимізацію узагальненого показника її ефективності.

Іл. 3. Бібліогр.: 10 назв.

UDC 621.396

Criteria and indexes of efficiency of the steganography systems of priv / *A.A. Smirnov* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 189 – 197.

The task of steganography priv, the primary purpose of which is a concealment not only of semantic maintenance of transferrable information but also the fact of realization of secretive transmission is examined. A mathematical model and flow diagram of the steganography system is probed. Basic operators are entered, legalistically describing basic calculable processes at stenographic transformation of information on the transmitting and receiving side. Criteria and indexes of efficiency of functioning of the steganography system are introduced. In a general view the task of perfection of steganosystems is formalized through maximization of the generalized index of its efficiency.

Fig.: 3. Ref.: 10 items.

УДК 621.391.7:336.71(075.8)

Сравнительный анализ протоколов строгой аутентификации / *И.В. Олешко, И.Д. Горбенко* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 198 – 209.

Оцениваются протоколы строгой аутентификации на основании условных и безусловных критериев. Показывается, что наилучшим протоколом аутентификации является протокол, основанный на сертификатах с использованием преобразований в группе точек эллиптической кривой.

Табл. 1. Ил. 7. Библиогр.: 9 назв.

УДК 621.391.7:336.71(075.8)

Порівняльний аналіз протоколів суворої автентифікації / *І.В. Олешко, І.Д. Горбенко* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 198 – 209.

Проводиться оцінка протоколів суворої автентифікації на основі умовних і безумовних критеріїв. Показано, що найкращим протоколом автентифікації є протокол, заснований на сертифікатах з використанням перетворень в групі точок еліптичної кривої.

Табл. 1. Іл. 7. Бібліогр.: 9 назв.

UDC 621.391.7:336.71(075.8)

Comparative analysis of strong authentication protocols / *I.V. Oleshko, I.D. Gorbenko* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 198 – 209.

The estimate of the strong authentication protocol based on conditional and unconditional criteria is carried out. It is shown that the best authentication protocol is the protocol based on the certificate with transformations in the group of the elliptic curve points.

Tab.1. Fig.7. Ref.:9 items.

РАДИОЛОКАЦИЯ
RADAR

УДК 621.396:621.391.82

Пространственная селекция широкополосных источников по собственным числам ковариационной матрицы / *О.В. Сытник, В.М. Карташов, А.А. Супрун* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С.210 – 215.

Рассматривается статистически оптимальный алгоритм пространственно-временной обработки сигналов для селекции точечных источников шумовых сигналов. В основу алгоритма положено представление пространственно-временной ковариационной матрицы наблюдения в виде набора матриц, полученных из усредненных сигналов многоотводных линий задержки приемных каналов адаптивной антенной решетки. Это позволило непосредственно по анализу собственных чисел строить оценки углового положения источников. Обработка сигнала при этом выполняется в несколько этапов, включающих оценивание пространственно-временной ковариационной матрицы по сигналам каналов приема, вычисления ее собственных чисел их ранжирования и оценивания соответствующих весовых коэффициентов матрицы усилителей. Выигрыш в чувствительности по отношению к клас-

ческому способу обработки в значительной мере зависит от допустимого интервала наблюдения, на котором осуществляется прием сигналов.

Библиогр.:10 назв.

УДК 621.396:621.391.82

Просторова селекція широкосмугових джерел за власними числами коваріаційної матриці / *О.В. Ситнік, В.М.Карташов, О.О. Супрун* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С.210 – 215.

Розглядається статистично оптимальний алгоритм просторово-часової обробки сигналів для селекції точкових джерел шумових сигналів. В основу алгоритму покладено уявлення про просторово-часову коваріаційну матрицю спостереження у вигляді набору матриць, отриманих з усереднених сигналів багатовідвідної ліній затримки приймальних каналів адаптивної антенної решітки. Це дозволило безпосередньо з аналізу власних чисел будувати оцінки кутового положення джерел. Обробка сигналу при цьому виконується в кілька етапів, що включають оцінювання просторово-часової коваріаційної матриці за сигналами каналів прийому, обчислення її власних чисел подальше їх ранжування та оцінювання відповідних вагових коефіцієнтів матриці підсилювачів. Виграш в чутливості по відношенню до класичного способу оброблення в значній мірі залежить від допустимого інтервалу спостереження, на якому здійснюється прийом сигналів.

Бібліогр.: 10 назв.

UDC 621.396:621.391.82

Spatial selection of wide-band sources using eigenvalue of the covariance matrix / *O.V. Sytnik, V.M.Kartashov, O.O. Suprun* // Radiotekhnika: All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 210 – 215.

The statistically optimal space-time signal processing algorithm for the point sources selection of noise signals is considered. The representation of the space-time covariance observation matrix as a set of matrices obtained from the averaged signals of the multipoint delay lines of the adaptive antenna array receiving channels is used as the algorithm basis. This makes it possible to construct the angular position estimates of the sources on the eigenvalue analysis. The signal processing in this case includes the space-time covariance matrix estimation by the signals of the receiving channels, the eigenvalues calculation of the space-time covariance matrix, the ranking calculation of eigenvalues, the corresponding coefficients estimation of the amplifier matrix. The gain in the sensitivity in relation to the classical processing method depends mainly on the tolerance observation interval during which the signals are received.

Ref.: 10 items.

УДК 621.396

Расчет характеристик обнаружения сигналов, отраженных объектами, расположенными на морской поверхности, полуактивной бистатической РЛС с цифровым телевизионным сигналом подсвета / *А.А. Байздренко, В.Б. Лубский* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 216 – 221.

Многочисленные исследования в области полуактивных бистатических РЛС с телевизионным подсветом подтвердили потенциальную возможность использовать телевизионный сигнал для радиолокационного обнаружения воздушных целей. Телевизионный сигнал по своим энергетическим характеристикам пригоден для обнаружения объектов также и на морской поверхности. Специфика распространения радиоволн над морской поверхностью требует разработки соответствующих методик расчетов характеристик обнаружения сигналов, отраженных морскими целями. Авторами статьи предлагается вариант расчетов характеристик обнаружения сигналов, отраженных объектами, расположенными на морской поверхности, бистатической РЛС с цифровым телевизионным сигналом подсвета.

Ил. 1. Библиогр.: 21 назв.

УДК 621.396

Розрахунок характеристик виявлення сигналів, відбитих об'єктами, розташованими на морській поверхні, полуактивною бістатичною РЛС з цифровим телевізійним сигналом освітлення / *О.О. Байздренко, В.Б. Лубський* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 216 – 221.

Численні дослідження у галузі полуактивних бістатичних РЛС з телевізійним освітленням підтвердили потенційну можливість застосовувати телевізійний сигнал для радіолокаційного виявлення повітряних цілей. Телевізійний сигнал за своїми енергетичними характеристиками придатний для виявлення об'єктів також і на морський поверхні. Специфіка розповсюдження радіохвиль над морською поверхнею вимагає розробку відповідних методик розрахунків характеристик виявлення сигналів, відбитих морськими цілями. Авторами статті пропонується варіант розрахунку характеристик вияв-

лення сигналів, відбитих об'єктами, розташованими на морський поверхні, бістатичною радіолокаційною станцією з цифровим телевізійним сигналом освітлення.

Іл. 1. Бібліогр.: 21 назв.

UDC 621.396

Calculation of detection characteristics of signals, reflected by objects located on the sea surface, in semi-active bistatic radar with digital TV signal illumination / *A.A. Baizdrenko, V.B. Lubsky* // Radiotekhnika: All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 216 – 221.

Numerous studies in the field of semi-active bistatic radars with television illumination confirmed the potential to use the TV signal for the radar detection of air targets. TV signal due to its energy performance is also suitable for detection of objects on the sea surface. Specificity of radio wave propagation over the sea surface requires the development of appropriate methods of calculation of the detection characteristics of the signals reflected by sea targets. A version of such calculations is presented.

1 fig. Ref.:21 items.

УДК 621.396.2

Сравнительный анализ помехозащищенности и спектральной эффективности Wi-Fi каналов связи с линейными и двумерными адаптивными антенными решетками при воздействии нескольких помех и одного сигнала / *А.А. Буланый, Г.В. Майстренко, А.А. Стрельницкий, В.М. Шокало* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 222 – 228.

Для частного случая прихода сигнала с направления близкого к нормали адаптивной антенной решетки (ААР) проведен сравнительный анализ помехозащищенности (уровень BER) и спектральной эффективности (СЭ) рэлеевских Wi-Fi каналов связи с линейными и двумерными ААР при воздействии нескольких помех и одного сигнала. Исследовались линейная и квадратная ААР, состоящие из четырех излучателей. Показано, что при воздействии одной помехи канал связи с квадратной решеткой по параметрам BER и СЭ существенно выигрывает по сравнению со случаем применения линейной ААР. Если же количество помех равно двум или трем, то более эффективным является канал с линейной решеткой. Определены значения отношения сигнал/помеха на входе ААР, при которых в канале связи может быть достигнуто высокое качество передачи информации.

Табл. 6. Ил. 4. Библиогр.: 6 назв.

УДК 621.396.2

Порівняльний аналіз завадозахищеності і спектральної ефективності Wi-Fi каналів зв'язку з лінійними та двовимірними адаптивними антенними решітками при впливі декількох завад і одного сигналу / *О.А. Буланий, Г.В. Майстренко, О.О. Стрельницький, В.М. Шокало* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 222 – 228.

Для частного випадку приходу сигналу з напрямку близького до нормалі адаптивної антенної решітки (ААР) проведено порівняльний аналіз завадозахищеності (рівень BER) і спектральної ефективності (СЕ) релєєвських Wi-Fi каналів зв'язку з лінійними та двовимірними адаптивними антенними решітками (ААР) при впливі декількох завад і одного сигналу. Досліджувалися лінійна і квадратна ААР, що складаються з чотирьох випромінювачів. Показано, що при впливі однієї завади канал зв'язку з квадратною решіткою по параметрам BER і СЕ виграє в порівнянні з випадком застосування лінійної ААР. Якщо ж кількість завад дорівнює двом або трьом, то ефективнішим є канал з лінійною решіткою. Визначено значення відносини сигнал/завада на вході ААР при яких в каналі зв'язку може бути досягнуто високу якість передачі інформації.

Табл. 6.Іл. 4. Бібліогр.: 6 назв.

UDC 621.396.2

Comparative analysis of immunity and spectral efficiency of Wi-Fi channels with linear and two-dimensional adaptive antenna array under the influence of several interferences and a signal / *A.A. Bulany, G.V. Maistrenko, A.A. Strelnitsky, V.M. Shokalo* // Radiotekhnika: All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 222 – 228.

A comparative analysis of noise immunity (BER level) and spectrum efficiency (SE) of Rayleigh Wi-Fi channels with linear and two-dimensional adaptive antenna arrays (AAA) under the influence of several interferences and a signal is conducted for the particular case when the signal arrives from the direction near to the normal of AAA. Linear and square AAA consisting of four emitters are studied. It is demonstrated that under one interference action the communication channel with a square lattice according to the BER and SE parameters benefits significantly as compared to the case of the linear AAA application. The channel with a

linear array is more efficient if the amount of interferences is two or three. The values of the signal/noise ratio at the AAA input, wherein a high quality of information transmission can be attained, are defined.

6 tab. 4 fig. Ref.: 6 items.

УДК 621.376

Исследование ФАР с диэлектрическим заполнениемем и согласующей периодической структурой / *С.В Марченко, В.М Морозов., А.М. Съянов* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 229 – 233.

Приведены результаты исследования бесконечной линейной волноводной ФАР для случая сканирования в Н-плоскости. Электродинамический алгоритм построен на основе метода пронизывающей области, который приводит к интегральному уравнению Фредгольма второго рода. Представлены численные результаты влияния параметров диэлектрического заполнения и размеров согласующей периодической структуры на согласование ФАР.

Ил. 6. Библиогр.: 4 назв.

УДК 621.376

Дослідження ФАР з діелектричним заповненням та узгоджуючою періодичною структурою / *С.В.Марченко, В.М Морозов, О.М. С'янов* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С.229 – 233.

Наведено результати дослідження нескінченої лінійної хвилеводної ФАР для випадку сканування в Н-площині. Електродинамічний алгоритм побудовано на основі методу області, що пронизує, який призводить до інтегрального рівняння Фредгольма другого роду. Представлено чисельні результати впливу параметрів діелектричного заповнення та розмірів узгоджуючої періодичної структури на узгодження ФАР.

Іл. 6. Бібліогр.: 4 назв.

UDC 621.376

Investigation of PAA with dielectrical inclusion and matching periodical structure/ *S.V. Marchenko, V.M Morozov., A.M Syanov* // Radiotekhnika: All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 229 – 233.

Results of investigation into the infinite linear waveguide PAA in the case of scanning in H-plane are presented. The electrodynamical algorithm is formulated on the basis of the penetrating area method which results in the second kind Fredholm equation. Numerical results of the dielectric inclusion and matching periodical structure parameters action on the PAA matching are presented.

6 fig. Ref.: 4 items.

УДК 551.501.8:621.396.96

Синтез и анализ дискриминатора следящего устройства систем радиоакустического зондирования атмосферы / *В.М. Карташов, Д.Н. Куля* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С.234 – 238.

Проанализированы особенности распространения радиоволн в условиях неоднородности атмосферы, искусственно вызванной акустическими колебаниями. Рассмотрены методы обработки отраженного радиосигнала от акустической посылки, позволяющие оценить температуру воздуха, и проанализированы причины погрешностей возникающих при обработке сигнала. Предложен корреляционный способ обработки, который обеспечивает высокую инструментальную точность регистрации вертикальных профилей температуры воздуха при сравнительно простом устройстве обработки рассеянных сигналов, где простота достигается главным образом за счет работы системы в режиме слежения за параметром расстройки условия Брэгга. Особенностью устройства является использование значения оцениваемого параметра, определенного на предыдущем цикле наблюдения.

Ил. 4. Библиогр.: 5 назв.

УДК 551.501.8:621.396.96

Синтез та аналіз дискримінатора слідкуючого пристрою систем радіоакустичного зондування атмосфери / *В.М. Карташов, Д.М. Куля* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 234 – 238.

Зроблено аналіз особливостей розповсюдження радіохвиль в умовах неоднорідності атмосфери, штучно викликаної акустичними коливаннями. Зроблено огляд методів обробки відбитого радіосигналу від акустичної посилки, що дозволяють оцінити температуру повітря та проаналізовані причини виникаючих похибок під час обробки сигналу. Запропоновано кореляційний спосіб обробки, котрий забезпечує високу інструментальну точність реєстрації вертикальних профілів температури повітря

при порівняно простому пристрої обробки розсіяних сигналів, де простота досягається за рахунок роботи системи в режимі слідкування за параметром розстройки умови Брега, особливістю котрого є використання значення параметра, визначеного на попередньому циклі спостереження.

Іл. 4. Бібліогр.: 5 назв.

UDC 551.501.8:621.396.96

Synthesis and analysis of the tracker device discriminator of the atmosphere radio acoustic sensing systems / *V.M.Kartashov, D.M.Kulia* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 234 – 238.

Features of the radio waves propagation in the atmosphere inhomogeneity induced by the acoustic vibrations are analyzed. Processing methods of the reflected radio signal from the speaker packages making it possible to estimate the air temperature are discussed. Causes of errors arising in the signal processing are analyzed. There is discussed the correlation processing method that provides high instrument accuracy of the vertical profiles registration of the air temperature at a relatively simple processing device of the scattered signals. Simplicity of the device is achieved by the operation of the system in the tracking mode for the parameter of the Bragg condition detuning. A feature of the device is to use the estimated parameter value defined in the previous cycle of observation.

4 fig. Ref.: 5 items.

УДК: 681.128.82

Контроль уровня легкоиспаряющихся жидкостей методом акустической локации / *Б.В. Жуков, А.В. Одновол* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 239 – 244.

Представлены результаты экспериментальных исследований функционирования акустических уровнемеров с абсолютным и относительным методами измерения в процессе контроля уровня нескольких типов жидкостей. Они показали, что на результаты контроля уровня легкоиспаряющихся жидкостей оказывают влияние их пары, которые локализуются непосредственно над поверхностью.

Табл. 1 Ил. 6. Библиогр.: 4 назв.

УДК: 681.128.82

Контроль рівня легковипаровуючих рідин методом акустичної локації / *Б.В. Жуков, А.В. Одновол* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 239 – 244.

Представлено результати експериментальних досліджень функціонування акустичних рівнемірів з абсолютним та відносним методами вимірювання в процесі контролю рівня декількох типів рідин. Вони показали, що на результати контролю рівня легковипаровуючих рідин впливають їх пари, які локалізуються безпосередньо над поверхнею.

Табл.1. Іл.6. Бібліогр.: 4 назв.

UDC 681.128.82

Control over the level of easily evaporated liquids by the acoustic location method / *B.V. Zhukov, A.V. Odnovol* // Radiotekhnika: All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P.239 – 244.

Results of experimental researches of acoustic level meters functioning with absolute and relative measurement methods in the process of several types of liquids' level monitoring are presented. They showed that the vapors of the easily evaporated liquids localized directly above their surface act on the results of their level control.

1 tab. 6 fig. Ref.: 4 items.

УДК 535.+543.47+621.396

Об аналитических методах синтеза поляризационного эллипса / *Г.М. Чекалин, Г.Н. Чекалина* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С.245 – 251.

Статья посвящена исследованию влияния исходной ориентации эллипса поляризации электромагнитной волны на вид аналитических выражений компонент поля и их разности фаз, а также коэффициента эллиптичности r и угла ориентации β при повороте системы отсчета на угол $\pm\beta$.

Ил. 3. Библиогр.: 3 назв.

УДК 535.+543.47+621.396

Про аналітичні методи синтезу поляризаційного еліпса / *Г.М. Чекалін, Г.М. Чекаліна* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 245 – 251.

Стаття присвячена дослідженню впливу вихідної орієнтації еліпса поляризації електромагнітної хвилі на вигляд аналітичних виразів компонент поля і їх різниці фаз, а також коефіцієнта еліптичності r і кута орієнтації β при повороті системи відліку на кут $\pm\beta$.

Іл. 3. Бібліогр.: 3.

UDC 535.+543.47+621.396

About analytical methods for synthesis of the polarization ellipse / *G.M. Chekalin, G.N. Chekalina* // Radiotekhnika: All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 245 – 251.

The action of the polarization ellipse initial orientation of the electromagnetic wave on the analytic expressions form of the field components and their phase difference, and the coefficient of ellipticity r and orientation angle β, when rotating the reference frame at the angle $\pm\beta$, is investigated.

3 fig. Ref.: 3/

УДК 004.056

Результаты экспериментального исследования телефонных радиозакладных уст ройств методом нелинейной локации / *Ю.В. Лыков* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 252 – 257.

Исследован метод нелинейной локации применительно к поиску радиозакладных устройств в проводных телефонных линиях. Разработаны требования к нелинейному локатору. Экспериментально получены нелинейные свойства некоторых моделей телефонных радиозакладных устройств. Показаны потенциальные возможности данного метода.

Ил. 7. Библиогр.: 4 назв.

УДК 004.056

Результати експериментального дослідження телефонних радіозакладних пристроїв методом нелінійної локації / *Ю.В. Ликов* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 252 – 257.

Досліджено метод нелінійної локації стосовно до пошуку радіозакладних пристроїв в проводових телефонних лініях. Розроблено вимоги до нелінійного локатора. Експериментально отримані нелінійні властивості деяких моделей телефонних радіозакладних пристроїв. Показані потенційні можливості даного методу.

Іл. 7. Бібліогр.: 4 назв.

UDC 004.056

Results of experimental study of telephone radio bugs using nonlinear locations method / *Y.V. Lykov* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 252 – 257.

The method of nonlinear location for search of radio bugs in wired telephone lines is studied. The requirements to the nonlinear locator are developed. The nonlinear properties of some phones' radio bugs are obtained experimentally. The potentialities of this method are shown.

7 fig. Ref.: 4 items.

УДК 551.501.7

Модельно-структурный анализ эхосигналов акустического зондирования атмосферы / *В.И. Леонидов* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 258 – 261.

Исследуются реализации акустических эхосигналов полученных на действующей станции акустического зондирования установленной в центре мегаполиса. Анализ проводится с помощью разрабатываемой методики модельно-структурного анализа, при которой моделируются эхосигналы, обусловленные структурными элементами турбулентного поля температуры.

Ил. 4. Библиогр.: 2 назв.

УДК 551.501.7

Модельно-структурний аналіз ехосигналів акустичного зондування атмосфери / *В.І. Леонідов* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 258 – 261.

Досліджуються реалізації акустичних ехосигналів отриманих на діючій станції акустичного зондування встановленої в центрі мегаполіса. Аналіз проводиться за допомогою розроблювальної методики модельно-структурного аналізу, при якому моделюються ехосигнали, обумовлені структурними елементами турбулентного поля температури.

Іл. 4. Бібліогр.: 2 назв.

UDC 551.501.7

Model-structural analysis of echo signals of atmospheric acoustic sounding / *V.I. Leonidov* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 258 – 261.

Realizations of the acoustic echo signals received at the working station of acoustic sounding and installed in the center of megalopolis are analyzed. The analysis is carried out using the structural-modeling analysis method, which is being presently developed and in accordance with which the echo signals stipulated by the structural elements of the temperatures turbulent field are modeled.

4 fig. Ref.: 2 items.

СИСТЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ
INFORMATION PROCESSING SYSTEMS

УДК 621.391

Разработка рекомендаций по регулированию пропускной способности в WPAN / *И.С. Шостко, Ю.Э. Соседки, Алмакадма Таха* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 262 – 269.

Для увеличения скорости передачи данных в радиоканале и обеспечения высокой помехоустойчивости в условиях многолучевого распространения и воздействия помех необходимо в перспективных системах WPAN использовать технологию UWB Multiband OFDM при оптимальном управлении параметрами сигнала. Для оценки показателей качества сверхширокополосных беспроводных систем связи, разработана модель приёмо-передатчика с использованием данной технологии. Модель позволяет исследовать: зависимости вероятности битовой ошибки от скорости свёрточного кодирования, от особенностей канала связи, от расстояния между приёмником и передатчиком и зависимость вероятности битовой ошибки от размера символа OFDM.

Ил. 10. Библиогр: 8 назв.

УДК 621.391

Розробка рекомендацій щодо регулювання пропускної здатності в WPAN / *І.С. Шостко, Ю.Е. Соседка, Алмакадма Таха* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 262 – 269.

Для збільшення швидкості передачі даних в радіоканалі і забезпечення високої завадостійкості в умовах багатопроменевого поширення і впливу перешкод необхідно в перспективних системах WPAN використовувати технологію UWB Multiband OFDM при оптимальному керуванні параметрами сигналу. Для оцінки показників якості надширокосмугових безпровідних систем зв'язку, розроблена модель прийомо-передавача з використанням даної технології. Модель дозволяє досліджувати: залежності ймовірності бітової помилки від швидкості згортального кодування, від особливостей каналу зв'язку, від відстані між приймачем і передавачем і залежність імовірності бітової помилки від розміру символу OFDM.

Іл. 10. Бібліогр.: 8 назв.

UDC 621.391

Development of recommendations concerning adjustment of carrying capacity in WPAN / *I.S. Shostka, J.E Sosedka, Almakadma Taha* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 262 – 269.

To increase the data transmission speed in the radio channel and to provide high immunity on the multipath dissemination condition and on the interference impact condition it is necessary in the perspective WPAN systems to use the UWB Multiband OFDM technology for the signal parameter optimum management. To assess the quality of UWB wireless communication systems, a model transceiver using this technology is developed. The model makes it possible to investigate the bit error probability dependence on the rate convolutional coding, the communication channel features, the distance between the receiver and the transmitter, the size of the OFDM symbol.

10 fig. Ref.: 8 items.

УДК 621.391.23

Особенности передачи данных по блокам в транкинговой системе стандарта APCO 25 / *И.В. Ковтун* // Радиотехника: Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 270 – 274.

Рассмотрена сравнительная характеристика основных перспективных стандартов цифровой транкинговой связи TETRA, APCO 25 и TETRAPOL по эксплуатационно-техническому и организационно – экономическому критериям. Проведена оценка объема передачи данных по блокам в стан-

дарте APCO 25. Рассчитано время передачи данных для блоков с двумя, тремя и четырьмя суперкадрами.

Табл. 1. Ил. 1.Библиогр.: 2 назв.

УДК 621.391.23

Особливості передачі даних по блоках у транкінгової системі стандарту APCO 25 / *І.В. Ковтун* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 270 – 274.

Розглянуто порівняльна характеристика основних перспективних стандартів цифрового транкінгового зв'язку TETRA, APCO 25 і TETRAPOL за експлуатаційно-технічним та організаційно-економічним критеріям. Проведено оцінку обсягу передачі даних по блоках у стандарті APCO 25. Розраховано час передачі даних для блоків з двома, трьома і чотирма суперкадрами.

Табл. 1. Іл. 1.Бібліогр.: 2 назви.

UDC 621.391.23

Especially data on blocks in the trunking system standard APCO 25 / *I.V. Kovtun* // Radiotekhnika: All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 270 – 274.

The comparative characteristics of the main perspective standards of the digital trunked communication TETRA, APCO 25 and TETRAPOL by the operational-technical and organizational-economic criteria are considered. The evaluation of the data transmission volume to the blocks in the APCO 25 standard is carried out. Time of data transmission to the blocks with two, three and four super-frames is calculated.

1 tab. 1 fig. Ref.: 2 items.

УДК 621.391.001

Помехоустойчивость m-позиционного автокорреляционного приемника шумовых сигналов в гауссовом канале / *Ю.Г.Лега* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 275 – 282.

Представлены результаты теоретического исследования помехоустойчивости m-позиционного автокорреляционного приемника шумовых сигналов в канале с аддитивным белым гауссовым шумом.

Ил. 4. Библиогр.: 3 назв.

УДК 621.391.001

Завадостійкість m-позиційного автокореляційного приймача шумових сигналів в гауссовому каналі / *Ю.Г.Лега* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 275 – 282.

Представлено результати теоретичного дослідження завадостійкості m-позиційного автокореляційного приймача шумових сигналів в каналі з адитивним білим гауссовим шумом.

Іл. 4. Бібліогр.: 3 назв.

UDC 621.391.001

Noise-immunity of m-positional autocorrelation receiver of noise signals in the Gaussian channel / *Yu. G. Lega* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 275 – 282.

The results of theoretical researches of noise-immunity m-positional autocorrelation receiver of noise signals in the channel with additive white Gaussian noise are presented.

4 fig. Ref.: 3 items.

УДК 621.397.7

Особенности межкадрового сжатия видеоинформации в устройствах видеонаблюдения и видеорегистрации / *С.А.Шейко* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 283 – 289.

Выявляются и анализируются причины появления «ложных» векторов движения при сжатии видеоизображений в системах видеонаблюдения и видеорегистрации. Показано, что использование оценочных функций, учитывающих битовые затраты на передачу векторов движения, позволяет избежать появления векторов при отсутствии движения, а также сократить скорость цифрового потока в системах видеонаблюдения до 10% в зависимости от степени сжатия. Также показана эффективность применения иерархического метода поиска векторов движения в видеопоследовательностях при малых отношениях сигнал-шум.

Ил. 5. Библиогр.: 8 назв.

УДК 621.397.7

Особливості міжкадрового стиснення відеоінформації в пристроях відеоспостереження та відеореєстрації / *С.О. Шейко* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 283 – 289.

Виявляються та аналізуються причини виникнення «помилкових» векторів руху при стисненні відеозображень у системах відеоспостереження та відеореєстрації. Показано, що використання оціночних функцій, які враховують бітові витрати на передачу векторів руху, дозволяє уникнути появи векторів при відсутності руху, а також скоротити швидкість цифрового потоку в системах відеоспостереження до 10% залежно від ступеня стиснення. Також показано ефективність застосування ієрархічного методу пошуку векторів руху у відеопослідовностях при малих відношеннях сигнал-шум.

Іл. 5. Бібліогр.: 8 назв.

UDC 621.397.7

Particularities of inter frame coding of video information in the video monitoring and video recording devices / *S. O. Sheiko* // Radiotekhnika: All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 283 – 289.

Reasons for "false" motion vectors appearance in the process of inter frame video coding in video monitoring and video recording devices are revealed and analyzed. It is shown that the use of the estimators taking into account the bit expenditures for the motion vectors transmission makes it possible to avoid appearance of vectors and to reduce digital flow velocity in the video observation systems up to 10% depending on the compression ratio. The efficiency of the hierarchical method for the motion vectors search in the video succession in the case of small SNR is also shown.

5 fig. Ref.: 8 items.

УДК 681.326

Методы планирования ресурсов в распределенных компьютерных системах / *В.Г. Котух, М.А. Мирошник, С.Н. Селевко* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 290 – 299.

Предложен подход и модель планирования вычислительных ресурсов в двухуровневой *GRID*-системе. Разработана динамическая процедура планирования распределения ресурсов в гетерогенной среде на основе решения задачи о наименьшем покрытии. Разработан программный продукт, реализующий имитационную дискретно-событийную модель планирования. Проведены вычислительные эксперименты на основе программной реализации модели, обосновывающие эффективность предложенной модели планирования распределением ресурсов в гетерогенных средах в выбранных метриках производительности работы системы. Показано, что предложенная процедура планирования позволяет существенно повысить равномерность загрузки гетерогенных ресурсов системы, уменьшить время выполнения всей очереди заданий в *GRID*-системе по сравнению с распространенным методом *FIFO*. Приведен вариант реализации предложенного метода в планировщике *MAUI*

Ил. 2. Библиогр.: 8 назв.

УДК 681.326

Методи планування ресурсів в розподілених комп'ютерних / *В.Г. Котух, М.А. Мірошник, С.М. Селевко* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 290 – 299.

Предложен подход и модель планирования вычислительных ресурсов в двухуровневой *GRID*-системе. Разработана динамическая процедура планирования распределения ресурсов в гетерогенной среде на основе решения задачи о наименьшем покрытии. Разработан программный продукт, реализующий имитационную дискретно-событийную модель планирования. Проведены вычислительные эксперименты на основе программной реализации модели, обосновывающие эффективность предложенной модели планирования распределением ресурсов в гетерогенных средах в выбранных метриках производительности работы системы. Показано, что предложенная процедура планирования позволяет существенно повысить равномерность загрузки гетерогенных ресурсов системы, уменьшить время выполнения всей очереди заданий в *GRID*-системе по сравнению с распространенным методом *FIFO*. Приведен вариант реализации предложенного метода в планировщике *MAUI*

Іл. 2. Бібліогр.: 8 назв.

UDC 681.326

Methods of resources planning in the distributed computer systems / *V.G. Kotuh, M.A. Miroshnik, S.N. Selevko* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 290 – 299.

The approach and model of scheduling of computing resources in two-level *GRID* is offered. Dynamic procedure of scheduling resources in the heterogeneous environment on the basis of the decision of a problem on the minimal cover is developed. The software product realizing the imitating discrete-event model of

scheduling is developed. Computing experiments on the basis of the models program realization proving efficiency of the offered model of scheduling resources in heterogeneous environments in chosen metrics of performance of work of system are carried out. It is shown that the offered procedure of planning allows raising essentially the uniformity of loading of the system heterogeneous resources, reducing the performance time for tasks in *GRID*-system in comparison with the widespread combined method on the basis of *FCFS* method.

2 fig. Ref.: 8 items.

РАДИОТЕХНИЧЕСКИЕ УСТРОЙСТВА
RADIO ENGINEERING DEVICES

УДК 621.372.852

Прямой метод синтеза полосно-пропускающих шлейфовых фильтров с чебышевской характеристикой / *Л.М. Карпуков, Р.Ю. Корольков* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 300 – 305.

Предложен метод синтеза полосно-пропускающих фильтров с симметричной структурой, составленной из четвертьволновых отрезков линий и короткозамкнутых шлейфов. Метод позволяет непосредственно по техническому заданию без использования фильтров-прототипов составлять функцию фильтрации, обеспечивающую равноволновую амплитудную характеристику фильтра, и определять волновые сопротивления элементов фильтра с помощью простой процедуры, основанной на перемножении матриц. Приведен пример синтеза фильтра, показывающий, что по предложенному методу в отличие от традиционного метода на основе фильтра-прототипа, получается строго равноволновая характеристика фильтра при значениях волновых сопротивлений его элементов в большей мере удовлетворяющих требованиям конструктивной реализации.

УДК 621.372.852

Прямий метод синтезу смуго-пропускаючих шлейфових фільтрів з чебишевською характеристикою / *Л.М. Карпуков, Р.Ю. Корольков* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 300 – 305.

Запропоновано метод синтезу смуго-пропускаючих фільтрів із симетричною структурою, складеної з чвертьхвильових відрізків ліній і короткозамкнутих шлейфів. Метод дозволяє безпосередньо за технічним завданням без використання фільтрів-прототипів складати функцію фільтрації, що забезпечує рівнохвильову амплітудну характеристику фільтра, і визначати хвильові опори елементів фільтра за допомогою простої процедури, яка заснована на перемножуванні матриць. Наведено приклад синтезу фільтра, який показує, що за запропонованим методом на відміну від традиційного методу на основі фільтра-прототипу, виходить строго рівнохвильова характеристика фільтра при значеннях хвильових опорів його елементів, що в більшій мірі задовольняють вимогам конструктивної реалізації.

UDC 621.372.852

Direct method for synthesis of bandpass stub filters with Chebyshev characteristic / *L.M. Karpukov, R.Y. Korolkov* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 300 – 305.

The method for synthesis of bandpass filters with a symmetric structure, made of the quarter-wave segments of lines and short circuited stubs is offer. This method allows to make the function of filtration, providing equal to the wave peak characteristic of filter, and to determine the impendences of filter members by means of simple procedure, based on multiplying of matrices directly on a requirement specification without the use of filters-prototypes. The filter synthesis example is cited, showing, that according to the offered method, unlike the traditional method based on the filter-prototype, the filter strictly equal to the wave characteristic with the values of its elements impedances conforming to the needs of the structural realization is obtained.

УДК 621.372(075)

Схемы замещения конденсаторов и катушек индуктивности / *П.Ф.Лебедев, В.П.Дробышев* // Радиотехника: Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 306 – 312.

Рассмотрены три варианта методик преобразования RL- и RC-цепей на базовой частоте. Показано, что двухэлементные цепи катушек индуктивности и конденсаторов могут быть эквивалентно представлены двумя схемами замещения в частотной области, в которой допустимо такое представление. При этом возможны два подхода к анализу: либо эквивалентность осуществляют по входному

сопротивлению, либо по входной проводимости. Для катушек индуктивности более логичен первый вариант, для конденсаторов – второй.

Ил.4. Библиогр. 3 назв.

УДК 621.372(075)

Схеми заміщення конденсаторів та котушок індуктивності / *П.Ф. Лебедев, В.П. Дробышев* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 306 – 312.

Розглянуто три варіанти методик перетворення RL- и RC-кіл на базовій частоті.Показано, що двоелементні кола котушок індуктивності та конденсаторів можуть бути еквівалентно представлені двома схемами заміщення в частотній галузі, в якій допустиме таке представлення. При цьому можливі два підходи до аналізу: або еквівалентність здійснюють за вхідним опором,або за вхідною провідністю. Для котушок індуктивності більш логічним є першій варіант, а для конденсаторів – другий.

Іл. 4. Бібліогр. 3 назв.

UDC 621.372(075)

Equivalent circuits of condensers and inductance coils / *P.F. Lebedev, V.P. Drobyshev* // Radiotekhnika: All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 306 – 312.

Three versions of techniques for transformation of RL – and RC – circuits on the base frequency are considered. It is shown, that two-element circuits of inductance coils and condensers can be equivalently presented by two equivalent circuits in the frequency area where such representation is admissible. Thus two approaches to the analysis are possible: either the equivalence is realized on the entrance resistance, or on the entrance conductivity. The first version is more logical for the inductance coils, the second version is more logical for the condensers.

4 fig. Ref.: 3 items.

УДК 621.396.677.71

Анализ энергетических характеристик поперечной щели в широкой стенке прямоугольного волновода с локальным диэлектрическим включением / *Д.Ю. Пенкин, Л.П. Яцук* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 313 – 321.

Численно исследованы диапазонные характеристики поперечной щели, прорезанной в широкой стенке прямоугольного волновода над локальным диэлектрическим включением, при различных геометрических параметрах структуры и значениях диэлектрической проницаемости вставки. Выявлены условия получения её коэффициента излучения близкого к единице, что, согласно теории волноводно-щелевых излучателей, недостижимо в случае однородного заполнения волновода. Показана возможность управления в широких пределах величинами энергетических параметров волноводно-щелевого элемента как при излучении в свободное полупространство, так и в другой прямоугольный волновод.

Ил. 6. Библиогр.: 4 назв.

УДК 621.396.677.71

Аналіз енергетичних характеристик поперечної щілини у широкій стінці прямокутного хвилеводу із локальним діелектричним вставленням / *Д.Ю. Пенкін, Л.П. Яцук* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 313 – 321.

Чисельно досліджені діапазонні характеристики поперечної щілини, що прорізана у широкій стінці прямокутного хвилеводу над локальним діелектричним вставленням, за умови різних геометричних параметрів структури та значеннях діелектричної проникності вставки. Визначено умови отримання її коефіцієнта випромінювання близького до одиниці, що, згідно теорії хвилеводно-щілинних випромінювачів, недосяжно у разі однорідного заповнення хвилеводу. Показано можливість управління у широких межах величинами енергетичних параметрів хвилеводно-щілинного елементу як при випромінюванні у вільний напівпростір, так і в інший прямокутний хвилевід.

Іл. 6. Бібліогр.: 4 назв.

UDC 621.396.677.71

Analysis of power characteristics of the transversal slot in a wide wall of the rectangular waveguide with a local dielectric insertion / *D.Yu. Penkin, L.P. Yatsuk* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 313 – 321.

Scattering parameters of a cross slot, cut in a broadside of rectangular waveguide and located over a dielectric insertion of a finite length, are numerically investigated as functions of dimensions of waveguide-slot radiator and a dielectric constant of slab. The conditions, under which the radiation coefficient tends to be unity, are determined (note that such a result is unachievable for a waveguide with homogeneous inner configuration according to fundamentals of waveguide-slot radiator theory). An ability to control widely the

scattering parameters of waveguide-slot radiator, which emits either in free space or to another rectangular waveguide structure, is demonstrated.

6 fig. Ref.: 4 items.

УДК 621.372.832

Излучение электромагнитных волн электрически длинной щелью с диэлектрическим заполнением в узкой стенке многомодового прямоугольного волновода / *С. Л. Бердник* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 322 – 326.

Методом Галеркина решена задача об излучении электромагнитных волн электрически длинной щелью с диэлектрическим заполнением в узкой стенке прямоугольного волновода, возбуждаемого высшими типами волн. Показана возможность реализации многочастотного излучателя на основе такой структуры при возбуждении волновода несколькими типами волн на разных частотах.

Ил. 4. Библиогр.: 6 назв.

УДК 621.372.832

Випромінювання електромагнітних хвиль електрично довгою щілиною з діелектричним заповненням у вузькій стінці багатомодового прямокутного хвилеводу / *С. Л. Бердник* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 322 – 326.

Методом Гальоркіна розв'язано задачу про випромінювання електромагнітних хвиль електрично довгою щілиною з діелектричним заповненням у вузькій стінці прямокутного хвилеводу, збуджуваного вищими типами хвиль. Показано можливість реалізації багаточастотного випромінювача на основі такої структури при збудженні хвилеводу декількома типами хвиль на різних частотах.

Іл. 4. Бібліогр.: 6 назв.

UDC 621.372.832

Radiation of electromagnetic waves by electrically long slot with dielectric filling in the narrow wall of the multimode rectangular waveguide / *S. L. Berdnik* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 322 – 326.

The problem of the radiation of electromagnetic waves by electrically long slot with dielectric filling in the narrow wall of rectangular waveguide excited by higher types of waves have been solved by the Galerkin's method. It is shown that the multi-frequency radiator based on this structure has been realized in the case of excitation of the waveguide by several modes.

4 fig. Ref.: 6 items.

УДК 681.7.068.4

Моделирование влияния структуры фотонно-кристаллических волокон на распределение модового поля и потери оптического сигнала в их соединениях / *А.И. Филипенко, О.В. Сычова* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 327 – 331.

Исследованы факторы потерь в соединениях между собой и с SMF волокнами нескольких моделей ФКВ с разной структурой поперечного сечения, путем моделирования получены их распределения поля основной моды, количественно оценено и проанализировано влияние оптико-геометрических отклонений, возникающих при соединении, на оптические потери.

Табл. 2. Ил. 6. Библиогр.: 5 назв.

УДК 681.7.068.4

Моделювання впливу структури фотонно-кристалічних волокон на розподіл модового поля та втрати оптичного сигналу в їх з'єднаннях / *О.І. Филипенко, О.В. Сичова* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 327 – 331.

Досліджено фактори втрат у з'єднаннях між собою й з SMF волокнами декількох моделей ФКВ із різною структурою поперечного перерізу, шляхом моделювання отримані їхні розподіли поля основної моди, кількісно оцінене й проаналізоване вплив оптико-геометричних відхилень, що виникають при з'єднанні, на оптичні втрати.

Табл. 2. Іл. 6. Бібліогр.: 5 назв.

UDC 681.7.068.4

Modeling of the photonic crystal fiber structure impact on the mode field distribution and on the optical signal loss in their connections / *A. I. Filipenko, O.V. Sychova* // Radiotekhnika: All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 327 – 331.

Loss factors in connections among themselves and with SMFs of several models PCFs with different structure of cross-section are researched, distributions of a basic mode field PCFs are received by modeling

and impact of the optic-geometrical deviations arising at connection on optical losses is quantitatively estimated and analyzed.

2 tab. 6 fig. Ref.: 5 items.

УДК 621.315.592

Исследование и оптимизация пленочных кремниевых аморфных фотопреобразователей на *p-i-n*-структурах / *Н.И. Слипченко, В.А. Письменецкий, А.В. Фролов, Н.В. Герасименко, М.Ю. Гуртовой, Е.С. Глушко, Т.Е. Стыценко* // Радиотехника : Всеукр. межвед. науч.-техн. сб. – 2012. – Вып. №171. – С. 332 – 339.

Исследованы и оптимизированы параметры тандемной пленочной кремниевой *p-i-n* гетероструктуры. Проведены расчеты *p-i-n* ячеек и экспериментальные исследования подсистемы солнечной станции, состоящей из панелей на основе аморфного кремния, которые подтвердили целесообразность их оптимизации.

Ил. 12. Библиогр.: 5 назв.

УДК 621.315.592

Дослідження і оптимізація плівкових кремнієвих аморфних фотоперетворювачів на *p-i-n*-структурах / *М.І. Сліпченко, В.О. Письменецький, А.В.Фролов, М.В. Герасименко, М.Ю. Гуртовий, О.С. Глушко, Т.Є. Стиценко* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. № 171. – С. 332 – 339.

Досліджено і оптимізовано параметри тандемної плівкової кремнієвої *p-i-n* гетероструктури. Проведено розрахунки *p-i-n* комірок та експериментальні дослідження підсистеми сонячної станції, що складається з панелей на основі аморфного кремнію, які підтвердили доцільність їх оптимізації.

Іл. 12. Бібліогр.: 5 назв.

UDC 621.315.592

Research and optimization of p-i-n structures based on the amorphous silicon films / *N.I. Slipchenko, V.A. Pismenetskiy, A.V. Frolov, N.V. Gerasimenko, M.Y. Gurtovoy, L.G. Glushko, T.E. Stytsenko* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. – N171. – P. 332 – 339.

Parameters for tandem p-i-n heterostructure consisting of silicon films were investigated and optimized. The calculations of p-i-n cells and experimental studies of the solar sub station consisting of panels based on the amorphous silicon confirmed the feasibility of their optimization.

12 fig. Ref.: 5 items.

УДК 528::629.783

Разработка и тестирование новых эффективных методов и алгоритмов обнаружения и устранения фазовых скачков статических и кинематических ГНСС–наблюдений / *А.А. Жалило.* // Радиотехника : Всеукр. межвед. науч.-техн. сборник. – 2012. – Вып. 171. С. 340 - 371.

Рассматривается задача определения потерь счета циклов в одночастотных и двухчастотных фазовых ГНСС–наблюдениях и их исправления. Описаны результаты разработки и тестирования новых эффективных методов и алгоритмов обнаружения, оценки и исправления фазовых скачков как статических, так и кинематических ГНСС–наблюдений. Рассмотрены два отличающихся метода обработки наблюдений. Первый из них реализует интервальную итерационную процедуру обнаружения и оценки (фиксации) скачков одинарных разностей фаз между парами спутников с последующей их идентификацией на отдельных трассах «спутники – приемник». Расширить возможности и повысить надежность устранения фазовых скачков одночастотных и двухчастотных кинематических наблюдений в сложных условиях (разрывы и пропуски наблюдений) позволяет применение другого предложенного метода. Он основан на использовании приращений наблюдений по времени и специальной процедуры совместного МНК–оценивания совокупности континуальных (приращений координат и расхождений часов) и дискретных параметров (циклических фазовых скачков).

Табл. 1. Ил. 4 Библиогр.: 16 назв.

УДК 528::629.783

Розробка і тестування нових ефективних методів і алгоритмів виявлення та усунення фазових стрибків статичних і кінематичних ГНСС–спостережень / *О.О. Жаліло* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. – 2012. – Вип. 171. С. 340 - 371.

Розглядається задача визначення втрат рахунку циклів в одночастотних і двочастотних фазових ГНСС–спостереженнях та їх виправлення. Описані результати розробки і тестування нових ефективних методів і алгоритмів виявлення, оцінки і виправлення фазових стрибків як статичних, так і кіне-

матичних ГНСС–спостережень. Розглянуто два методи обробки спостережень, що відрізняються. Перший з них реалізує інтервальну ітераційну процедуру виявлення і оцінки (фіксації) стрибків одинарних різниць фаз між парами супутників з наступною їх ідентифікацією на окремих трасах «супутники – приймач». Розширити можливості і підвищити надійність усунення фазових стрибків одночастотних і двочастотних кінематичних спостережень в складних умовах (розриви і пропуски спостережень) дозволяє застосування іншого запропонованого методу. Він заснований на використанні приростів спостережень за часом і спеціальної процедури спільного МНК–оцінювання сукупності континуальних (приростів координат і розбіжностей годинників) і дискретних параметрів (циклічних фазових стрибків).

Табл.1. Іл. 4 . Бібліогр.: 16 назв.

UDC УДК 528::629.783

Development and testing of new methods and algorithms for detecting and correction the carrier-phase slips of static and kinematic GNSS observations / *A.A.Zhalilo* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. – 2012. № 171. – P. 340 - 371.

The task of determining cycle counting losses in single- and dual-frequency carrier-phase GNSS observations and their repair is studied. It is described the results of development and testing of new effective methods and algorithms for detection, assessment and repair of carrier-phase jumps (slips) of both static and kinematic GNSS observations. Two different methods of observation processing are considered. The first one implements interval iterative detection procedure and estimation (fixing) of slips of carrier-phase single differences between pairs of satellites with their subsequent identification on separate tracks «satellites–receiver». Using of other method enables to enhance possibilities and increase reliability of eliminating the carrier-phase slips of single- and dual-frequency kinematic observations in difficult circumstances (gaps and missing observations). It is based on using the increments of observations by time and special procedure for joint LSM–estimation of combining the continuum (increments of coordinates and clock variations) and discrete parameters (carrier-phase cycle slips).

1 Tab. 4 Fig. Ref.: 16 items.