MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE

KHARKIV NATIONAL
UNIVERSITY OF RADIO ELECTRONICS

RADIOTEKHNIKA

All-Ukrainian
interdepartmental scientific and technical collection

ISSN 0485-8972
eISSN 2786-5525

Founded in 1965

I S S U E 2 1 7

Kharkiv
Kharkiv National
University of Radio Electronics
2024

UDC 621.3

The collection is included in the List of scientific professional publications of Ukraine, category «Б», technical and physical-mathematical sciences (approved by orders of the Ministry of Education and Science from 17.03.2020 № 409; from 02.07.2020 № 886; from 24.09.2020 № 1188) by specialties: 105 – Applied Physics and Nanomaterials; 125 – Cybersecurity and information protection; 163 – Biomedical Engineering; 171 – Electronics; 172 – Telecommunications and Radio Engineering; 173 – Avionics; 174 –Automation and Computer-Integrated Technologies and Robotics; 175 – Metrology and information-measuring technique; 176 – Micro- and Nanosystem Technology.

Website: rt.nure.ua

Registration certificate KV № 12098-969 PR dated 14. 12. 2006.

The authors are responsible for the content of the article.

Responsible for the issue: *I.V. Svyd, PhD, Assoc. prof., I.D. Gorbenko, Dr. Sc. (Tech.), prof.*
Technical Secretary: *O.S. Polyakova.*

Recommended by the Scientific and Technical Council of Kharkiv National University of Radio Electronics, protocol № 5 dated 14.06.2024.

Address of the editorial board: Kharkiv National University of Radio Electronics (NURE), ave. Nauky, 14, Kharkiv, 61166, tel. (0572) 7021-397.

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ
УНІВЕРСИТЕТ РАДІОЕЛЕКТРОНІКИ

РАДІОТЕХНІКА

Всеукраїнський
міжвідомчий науково-технічний збірник

ISSN 0485-8972
eISSN 2786-5525

Засновано в 1965 р.

В И П У С К 2 1 7

Харків
Харківський національний
університет радіоелектроніки
2024

УДК 621.3

Збірник включено до Переліку наукових фахових видань України, категорія "Б", технічні та фізико-математичні науки (затверджено наказами МОНУ від 17.03.2020 № 409; від 02.07.2020 № 886; від 24.09.2020 № 1188) за спеціальностями: 105 – Прикладна фізика та наноматеріали; 125 – Кібербезпека та захист інформації; 163 – Біомедична інженерія; 171 – Електроніка; 172 – Телекомунікації та радіотехніка; 173 – Авіоніка; 174 – Автоматизація, комп'ютерно-інтегровані технології та робототехніка; 175 – Метрологія та інформаційно-вимірювальні технології; 176 – Мікро- та наносистемна техніка.

Сайт: rt.nure.ua

Реєстраційне свідоцтво КВ № 12098-969 ПР від 14. 12. 2006.

За зміст статті відповідальні автори.

Відповідальні за випуск: *І.В. Свид, канд. техн. наук, доц.*, *І.Д. Горбенко, д-р техн. наук, проф.*
Технічний секретар: *О.С. Полякова.*

Рекомендовано Науково-технічною радою Харківського національного університету радіоелектроніки, протокол № 5 від 14.06.2024.

Адреса редакційної колегії: Харківський національний університет радіоелектроніки (ХНУРЕ), просп. Науки, 14, Харків, 61166, тел. (0572) 7021-397.

CONTENT

SYSTEMS AND METHODS OF INFORMATION PROTECTION

MEANS OF TELECOMMUNICATIONS

RADIO ENGINEERING DEVICES

PHYSICS OF DEVICES, ELEMENTS AND SYSTEMS

RADIO ELECTRONIC SYSTEMS

ЗМІСТ

СИСТЕМИ І МЕТОДИ ЗАХИСТУ ІНФОРМАЦІЇ

ЗАСОБИ ТЕЛЕКОМУНІКАЦІЙ

РАДІОТЕХНІЧНІ ПРИСТРОЇ

ФІЗИКА ПРИЛАДІВ, ЕЛЕМЕНТІВ І СИСТЕМ

РАДІОЕЛЕКТРОННІ СИСТЕМИ

УДК 004.056 **DOI:10.30837/rt.2024.2.217.01**

О.В. ПОТІЙ, д-р техн. наук, Д.Ю. ГОЛУБНИЧИЙ, канд. техн. наук,
Ю.К. ВАСИЛЬЄВ, М.В. ЄСІНА, канд. техн. наук

ПРОЦЕС ДЕКЛАРУВАННЯ ПРОФІЛІВ БЕЗПЕКИ ІНФОРМАЦІЇ

Вступ

Нові виклики у сфері захисту інформації та кібербезпеки вимагають від держави та власників інформаційних активів застосовувати нові технології захисту з метою підвищення ефективності впровадження та використання засобів захисту. Тому на сьогодні з'явилась нагальна потреба у модернізації вітчизняної нормативної бази у сфері захисту інформації. Модернізація нормативної бази має проводитися у напрямку гармонізації з міжнародними стандартами, а також враховувати досвід світових лідерів у цій галузі – ЄС та США.

По суті, дана робота розглядає процес реалізації експериментального проєкту з декларування відповідності комплексних систем захисту інформації (КСЗІ) в інформаційних (ІС), електронних комунікаційних та інформаційно-комунікаційних системах (ІКС), створених з використанням профілів безпеки інформації. Особливістю подання матеріалу є застосування процесного підходу щодо розкриття нормативно-правових документів, що регламентують проведення експериментального проєкту з декларування відповідності КСЗІ з використанням профілів безпеки інформації.

Вважаємо, що актуальними та необхідними питаннями, які розглянуті в статті, будуть питання надання навчально-методичної допомоги власникам (розпорядникам) систем, які в умовах ведення війни з російською федерацією готові взяти на себе відповідальність за ефективність реалізованих заходів захисту ІС, ІКС, що знаходяться під їхньою експлуатацією.

1. Аналіз стандартів, на яких побудовані системи захисту інформації в інформаційних, електронних комунікаційних та інформаційно-комунікаційних системах

На сьогодні для забезпечення безпеки інформації, що оброблюється у системі, зазвичай застосовуються комплексні системи захисту інформації (КСЗІ), системи управління інформаційною безпекою (СУІБ) [1] та системи безпеки інформації (СБІ) [2]. Також можна використовувати вимоги та рекомендації до забезпечення безпеки інформації, що надає NIST SP 800-53 [3]. Також розроблений інший підхід до захисту інформації, що засновується на використанні базових профілів безпеки (БПБ) [4].

Безумовно, кожен з цих підходів до забезпечення безпеки інформації має свої переваги та недоліки. Проведені авторами статті дослідження дозволили проаналізувати та зробити висновки щодо переваг, недоліків та можливостей застосування кожного з підходів до забезпечення безпеки інформації.

Створення КСЗІ є важливим кроком для забезпечення безпеки даних та відповідності законодавству, але потребує значних ресурсів та кваліфікації для ефективного впровадження та підтримки.

Створення КСЗІ має свої переваги та недоліки, які варто враховувати під час прийняття рішення щодо їх впровадження [5].

До переваг можна віднести наступне:

1. Підвищений рівень безпеки: КСЗІ забезпечують багаторівневий захист від різноманітних загроз, таких як віруси, атаки хакерів, витоки інформації та інші кіберзагрози.

2. Відповідність законодавству: впровадження КСЗІ допомагає організаціям відповідати вимогам національного та міжнародного законодавства щодо захисту інформації.

3. Захист конфіденційних даних: КСЗІ дозволяють забезпечити конфіденційність і цілісність критично важливої інформації, що знижує ризики втрати або викрадення даних.

4. Безперервність бізнесу: завдяки захищеній інфраструктурі зменшуються ризики простоїв і втрати даних, що сприяє безперервності та стабільності бізнес-процесів.

5. Підвищення довіри клієнтів і партнерів: наявність сертифікованої системи захисту інформації підвищує довіру клієнтів і партнерів, що може бути конкурентною перевагою на ринку.

До недоліків можна віднести:

1. Високу вартість впровадження та підтримки: розробка, впровадження та підтримка КСЗІ можуть бути досить витратними, особливо для малих та середніх підприємств.

2. Складність управління: управління КСЗІ вимагає високого рівня кваліфікації персоналу, що може бути викликом для компаній, які не мають відповідних ресурсів.

3. Зниження продуктивності: деякі засоби захисту можуть впливати на продуктивність систем, уповільнюючи роботу користувачів і процесів.

4. Необхідність регулярного оновлення: КСЗІ потребують регулярного оновлення для захисту від нових загроз, що вимагає додаткових витрат та зусиль.

5. Потенційні проблеми з інтеграцією: інтеграція КСЗІ з існуючими системами може бути складною і вимагати значних зусиль для забезпечення сумісності та коректної роботи.

Створення та застосування СУІБ є важливим для забезпечення комплексного захисту інформації та відповідності нормативним вимогам. Проте, процес впровадження та підтримки таких систем може бути складним та затратним, що вимагає ретельного планування та наявності відповідних ресурсів [1].

СУІБ є важливими для захисту даних та забезпечення безпеки інформаційних активів організацій. Розглянемо переваги та недоліки створення та застосування таких систем.

Переваги:

1. Системний підхід до безпеки: СУІБ дозволяють систематизувати та стандартизувати процеси управління безпекою інформації, що сприяє більш ефективному виявленню та реагуванню на загрози.

2. Відповідність стандартам та нормативним документам: впровадження СУІБ допомагає організаціям відповідати міжнародним стандартам, наприклад, таким як ISO/IEC 27001, та вимогам місцевого законодавства, що може бути критичним для багатьох галузей.

3. Зниження ризиків: СУІБ дозволяють ідентифікувати, оцінювати та управляти ризиками, що знижує ймовірність виникнення інцидентів інформаційної безпеки.

4. Покращення репутації: наявність сертифікованої СУІБ підвищує довіру клієнтів, партнерів та акціонерів до компанії, покращуючи її імідж на ринку.

5. Підвищення ефективності бізнес-процесів: інтеграція СУІБ може сприяти оптимізації бізнес-процесів та покращенню управління інформаційними активами.

6. Забезпечення безперервності бізнесу: СУІБ допомагають забезпечити стійкість організації до інцидентів та аварійних ситуацій, що сприяє безперервності бізнесу.

Недоліки:

1. Висока вартість впровадження та підтримки: створення та підтримка СУІБ можуть бути дорогими, що може бути суттєвим викликом для малих та середніх підприємств.

2. Складність впровадження: процес впровадження СУІБ може бути складним і вимагати значних зусиль та часу для адаптації існуючих процесів та навчання персоналу.

3. Необхідність постійного моніторингу та оновлення: СУІБ потребують регулярного моніторингу, аудиту та оновлення для забезпечення актуальності та ефективності, що потребує додаткових ресурсів.

4. Складність управління змінами: впровадження нових політик та процедур може зустріти опір з боку персоналу, що потребує додаткових зусиль для управління змінами та навчання.

5. Зниження гнучкості: деякі компанії можуть виявити, що суворе дотримання політик та процедур СУІБ знижує гнучкість і швидкість реагування на нові бізнес-можливості.

6. Високі вимоги до кваліфікації персоналу: ефективне управління СУІБ вимагає високого рівня кваліфікації та спеціалізованих знань, що може бути викликом для деяких організацій.

Створення та застосування СБІ відповідно до НД ТЗІ 3.6-004-21 є важливим для забезпечення захисту інформації та відповідності законодавчим вимогам. СБІ передбачає комплекс організаційних, технічних, програмних та інших заходів для захисту інформації. Процес впровадження та підтримки таких систем може бути складним та затратним, що потребує ретельного планування та наявності відповідних ресурсів [2].

Розглянемо переваги та недоліки створення та застосування такої системи.

Переваги:

1. Відповідність законодавчим вимогам: впровадження СБІ дозволяє організаціям відповідати вимогам національного законодавства щодо захисту інформації, що є обов'язковим для державних та деяких приватних структур.

2. Комплексний підхід до безпеки: НД ТЗІ 3.6-004-21 описує комплекс заходів, що включають організаційні, технічні, програмні та інші аспекти безпеки, забезпечуючи всебічний захист інформації.

3. Систематизація процесів безпеки: впровадження СБІ дозволяє систематизувати процеси безпеки інформації, що сприяє більш ефективному управлінню ризиками та реагуванню на загрози.

4. Підвищення довіри клієнтів та партнерів: наявність сертифікованої СБІ підвищує довіру до організації з боку клієнтів, партнерів та інших зацікавлених сторін.

5. Захист від різноманітних загроз: НД ТЗІ 3.6-004-21 включає заходи для захисту від різноманітних загроз, включаючи кібератаки, витоки інформації та ін.

6. Підвищення загального рівня безпеки: впровадження заходів, описаних у документі, допомагає підвищити загальний рівень інформаційної безпеки в організації, знижуючи ризики втрат інформації.

Недоліки:

1. Висока вартість впровадження: комплексний підхід до безпеки може вимагати значних фінансових витрат на закупівлю обладнання, програмного забезпечення та навчання персоналу.

2. Складність впровадження: впровадження СБІ може бути складним процесом, що вимагає значного часу та ресурсів, особливо для великих організацій.

3. Необхідність постійного моніторингу та оновлення: СБІ потребує регулярного моніторингу, аудиту та оновлення для забезпечення актуальності та відповідності новим загрозам, що вимагає додаткових ресурсів.

4. Зниження гнучкості бізнес-процесів: впровадження суворих заходів безпеки може знижувати гнучкість бізнес-процесів, уповільнюючи реакцію на зміни ринкових умов або нові бізнес-можливості.

5. Високі вимоги до кваліфікації персоналу: управління та підтримка СБІ вимагають високого рівня кваліфікації та спеціалізованих знань, що може бути проблемою для деяких організацій.

6. Потенційні проблеми з інтеграцією: інтеграція нових систем безпеки з існуючими інформаційними системами може бути складною і вимагати додаткових зусиль для забезпечення сумісності та коректної роботи.

Застосування та реалізація вимог NIST SP 800-53 rev. 5 можуть значно підвищити рівень безпеки інформаційних систем і управління ризиками в організації. NIST SP 800-53 rev. 5 є важливим документом для забезпечення безпеки інформаційних систем і мереж. Він описує контрольні заходи та процедури для оцінки ефективності систем управління безпекою

інформації. Але його процес впровадження може бути складним і витратним, що потребує ретельного планування, наявності кваліфікованого персоналу та достатніх ресурсів [3].

Розглянемо переваги та недоліки застосування та реалізації вимог цього документа на практиці.

Переваги:

1. Комплексний підхід до безпеки: документ охоплює широкий спектр контрольних заходів і процедур для забезпечення всебічного захисту інформаційних систем, включаючи технічні, організаційні, адміністративні та фізичні аспекти.

2. Відповідність міжнародним стандартам: NIST SP 800-53 rev. 5 базується на міжнародно визнаних практиках і стандартах, що допомагає організаціям відповідати вимогам міжнародного ринку та регуляторів.

3. Гнучкість і адаптивність: документ пропонує гнучкий підхід до впровадження контрольних заходів, що дозволяє адаптувати їх до специфічних потреб і умов кожної організації.

4. Підвищення рівня безпеки: виконання вимог NIST SP 800-53 rev. 5 підвищує загальний рівень безпеки інформаційних систем, знижуючи ризики інцидентів безпеки та витоку даних.

5. Покращення процесів управління ризиками: документ допомагає організаціям краще ідентифікувати, оцінювати та управляти ризиками, що сприяє більш ефективному захисту інформаційних активів.

6. Підвищення довіри клієнтів і партнерів: дотримання вимог NIST SP 800-53 rev. 5 підвищує довіру клієнтів і партнерів до організації, покращуючи її репутацію та конкурентоспроможність на ринку.

Недоліки:

1. Високі витрати на впровадження: реалізація вимог документа може бути дорогим процесом, що включає закупівлю обладнання, програмного забезпечення та навчання персоналу.

2. Складність впровадження: впровадження контрольних заходів NIST SP 800-53 rev. 5 може бути складним процесом, що вимагає значних зусиль і ресурсів, особливо для організацій з обмеженими можливостями.

3. Високі вимоги до кваліфікації персоналу: ефективне впровадження та підтримка вимог документа вимагають високого рівня кваліфікації та спеціалізованих знань від персоналу.

4. Часові витрати: процес впровадження вимог NIST SP 800-53 rev. 5 може зайняти багато часу, що може впливати на загальну продуктивність організації.

5. Потенційна бюрократизація процесів: виконання детальних вимог і процедур може призвести до збільшення бюрократичних процесів в організації, що може уповільнити прийняття рішень і реалізацію проєктів.

6. Необхідність постійного моніторингу та оновлення: вимоги документа потребують регулярного моніторингу та оновлення для забезпечення актуальності та відповідності новим загрозам, що вимагає додаткових ресурсів.

2. Нормативно-правові акти щодо декларування відповідності комплексних систем захисту інформації в інформаційних, електронних комунікаційних та інформаційно-комунікаційних системах, створених з використанням профілів безпеки інформації

Нормативні документи (НД) системи технічного захисту інформації (ТЗІ), на яких ґрунтується КСЗІ, визначають процес створення певного підґрунтя для власників ІС, електронних комунікаційних та ІКС. Така система забезпечує власників ІС, електронних комунікаційних та ІКС чітким механізмом та допомагає в його реалізації. Вихідним документом, що регламентує порядок впровадження експериментального проєкту з декларування є Постанова Кабінету Міністрів України від 30.05.2024 № 627 (рис. 1) [4].

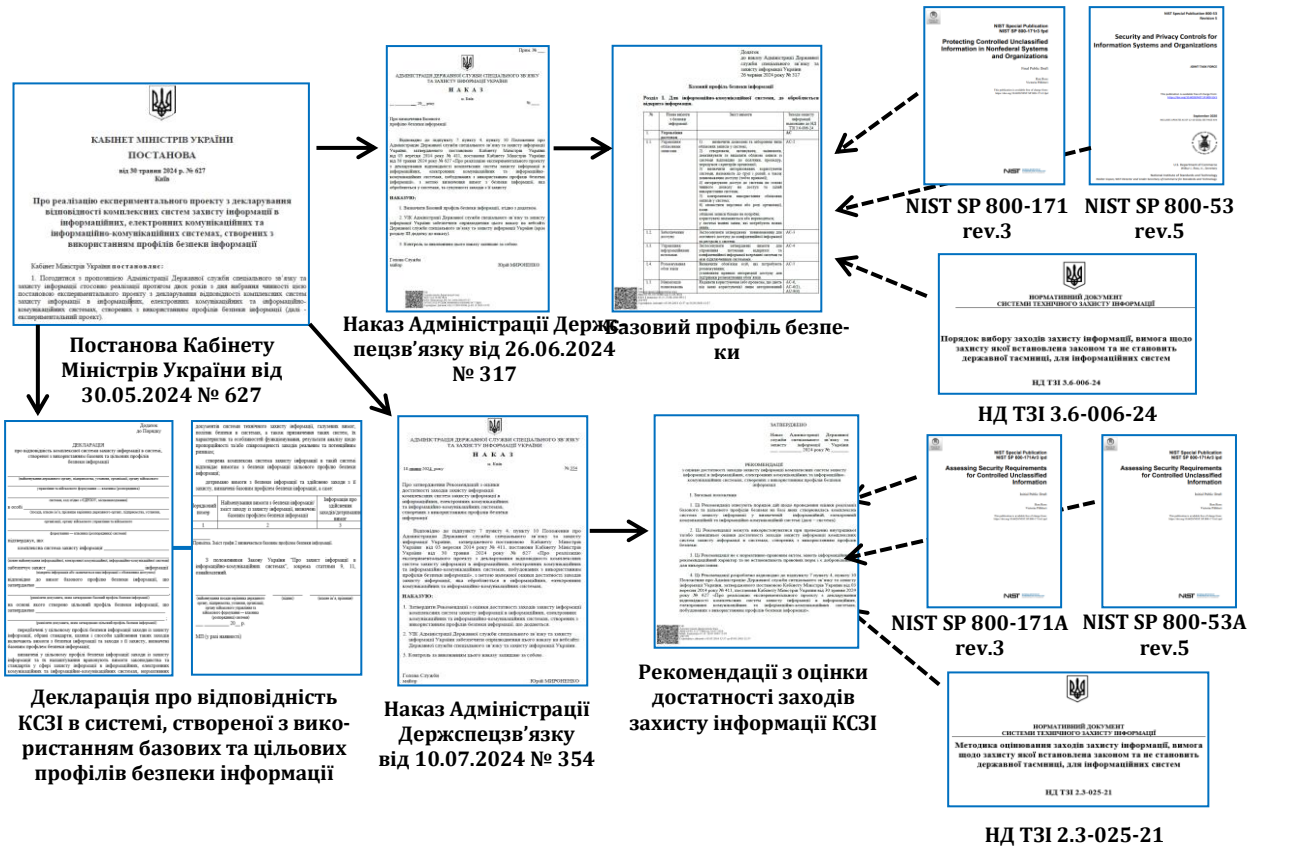


Рис. 1. Нормативно-правові акти щодо реалізації проєкту з декларування відповідності КСЗІ в інформаційних, електронних комунікаційних та інформаційно-комунікаційних системах, створених з використанням профілів безпеки інформації

Метою експериментального проєкту є підвищення рівня захисту державних інформаційних ресурсів (ДІР) та інформації, вимога щодо захисту якої встановлена законом, оптимізація процесу державної експертизи та підтвердження відповідності комплексних систем захисту інформації в системах.

Новий проєкт дозволить децентралізувати та спростити процеси оцінки та впровадження КСЗІ, підвищити відповідальність власника (розпорядника) системи. Нормативно-правові акти, що розроблені в рамках проєкту, враховують положення:

- посібника із заходів безпеки та приватності для інформаційних систем і організацій (NIST SP 800-53 rev. 5: Security and Privacy Controls for Information Systems and Organizations [3]), виданого Національним інститутом стандартів та технології Сполучених Штатів Америки (NIST);
- методології та набору процедур для проведення оцінки засобів захисту та конфіденційності, які застосовуються в системах і організаціях у рамках ефективного управління ризиками (NIST SP 800-53A rev. 5: Assessing Security and Privacy Controls in Information Systems and Organizations [6]);
- посібника із захисту контрольованої несекретної інформації в нефедеральних системах і організаціях (NIST SP 800-171 rev. 3. Protecting Controlled Unclassified Information in Nonfederal Systems and Organizations [7]);
- процедури оцінки та методології, які можна використовувати для проведення оцінки вимог безпеки в спеціальній публікації NIST 800-171 (NIST SP 800-171A rev. 3. Assessing Security Requirements for Controlled Unclassified Information [8]);
- порядку вибору заходів захисту інформації, вимога щодо захисту якої встановлена законом та не становить державної таємниці, для інформаційних систем (НД ТЗІ 3.6-006-24 [9]);

- методики оцінювання заходів захисту інформації, вимога щодо захисту якої встановлена законом та не становить державної таємниці, для інформаційних систем (НД ТЗІ 2.3-025-24 [10]).

Основним поняттям, на якому побудовані вказані нормативно-правові документи є поняття профілю безпеки. Профіль безпеки (ПБ) – набір заходів захисту, які застосовуються до інформації або ІС для задоволення вимог чинної нормативної бази, а також спрямовані на захист потреб з метою управління ризиками безпеки [2].

Створення комплексних систем захисту інформації в системах з використанням базових та цільових профілів та декларування відповідності таких комплексних систем здійснюється за рахунок налаштування базового профілю безпеки (БПБ).

Базовим профілем безпеки інформації є вимоги з безпеки інформації та взаємопов'язана сукупність заходів з її захисту, визначені Адміністрацією Держспецзв'язку для відкритої інформації та інформації з обмеженим доступом, яка обробляється у системах.

Базовий профіль безпеки є простим і доступним варіантом для організацій з обмеженими ресурсами та невисокими вимогами до безпеки. Він забезпечує мінімальний рівень захисту та може бути швидко впроваджений. Однак для організацій з високими вимогами до безпеки або специфічними загрозами, більш комплексні системи, такі як КСЗІ, СУІБ або СБІ, можуть бути більш відповідними, хоча і потребують значних ресурсів для впровадження та підтримки.

Цільовий профіль безпеки (ЦПБ) – представляє собою взаємопов'язану сукупність заходів із захисту інформації та їх налаштування, визначених для системи її власником (розпорядником) відповідно до базового профілю з урахуванням вимог законодавства та стандартів у сфері захисту інформації в ІС, електронних комунікаційних та ІКС, нормативних документів системи технічного захисту інформації, галузевих вимог, політик безпеки в системах, а також призначення системи, її характеристик та особливостей функціонування, результатів проведеної оцінки ризиків [4].

БПБ є корисним інструментом для встановлення мінімальних стандартів захисту інформації в організації. Він забезпечує швидке впровадження та зниження ризиків від найпоширеніших загроз, проте для повного захисту інформаційних активів може вимагати подальшого посилення та адаптації до специфічних потреб організації.

Основні характеристики базового профілю безпеки:

1. Мінімальні вимоги до безпеки: визначає мінімальний набір заходів, які повинні бути впроваджені для забезпечення базового рівня безпеки.

2. Стандартні контрольні заходи: включає стандартні контрольні заходи, такі як управління доступом, шифрування даних, безпека мережі, моніторинг та аудит, захист від шкідливого програмного забезпечення.

3. Універсальність: може бути застосований до різних типів організацій та інформаційних систем, забезпечуючи базовий рівень захисту незалежно від специфіки.

4. Спрощене впровадження: базовий профіль безпеки розроблений таким чином, щоб бути легким для впровадження, навіть для організацій з обмеженими ресурсами та технічними можливостями.

Переваги:

1. Швидке впровадження: набір мінімальних вимог дозволяє швидко розгорнути базові заходи безпеки, забезпечуючи початковий рівень захисту.

2. Зниження ризиків: реалізація БПБ знижує ризики від найпоширеніших загроз та вразливостей, забезпечуючи фундаментальний рівень захисту інформації.

3. Підвищення обізнаності: допомагає підвищити обізнаність організацій щодо необхідності забезпечення інформаційної безпеки та встановлення основних стандартів захисту.

Недоліки:

1. Обмежений рівень захисту: БПБ забезпечує лише мінімальний рівень захисту, що може бути недостатнім для організацій з високими вимогами до безпеки.

2. Необхідність подальшого розширення: організації, які впровадили БПБ, можуть потребувати подальшого розширення та посилення заходів безпеки для забезпечення захисту від більш складних та специфічних загроз.

Під час визначення цільового профілю власник (розпорядник) системи самостійно обирає стандарти у сфері захисту інформації в ІС, електронних комунікаційних та ІКС, які використовуються під час здійснення заходів із захисту інформації, шляхи і способи здійснення таких заходів відповідно до цільового профілю, а також визначає наявність у ньому інформації з обмеженим доступом та забезпечує дотримання встановлених правил роботи з документами, які містять інформацію з обмеженим доступом (рис. 2).

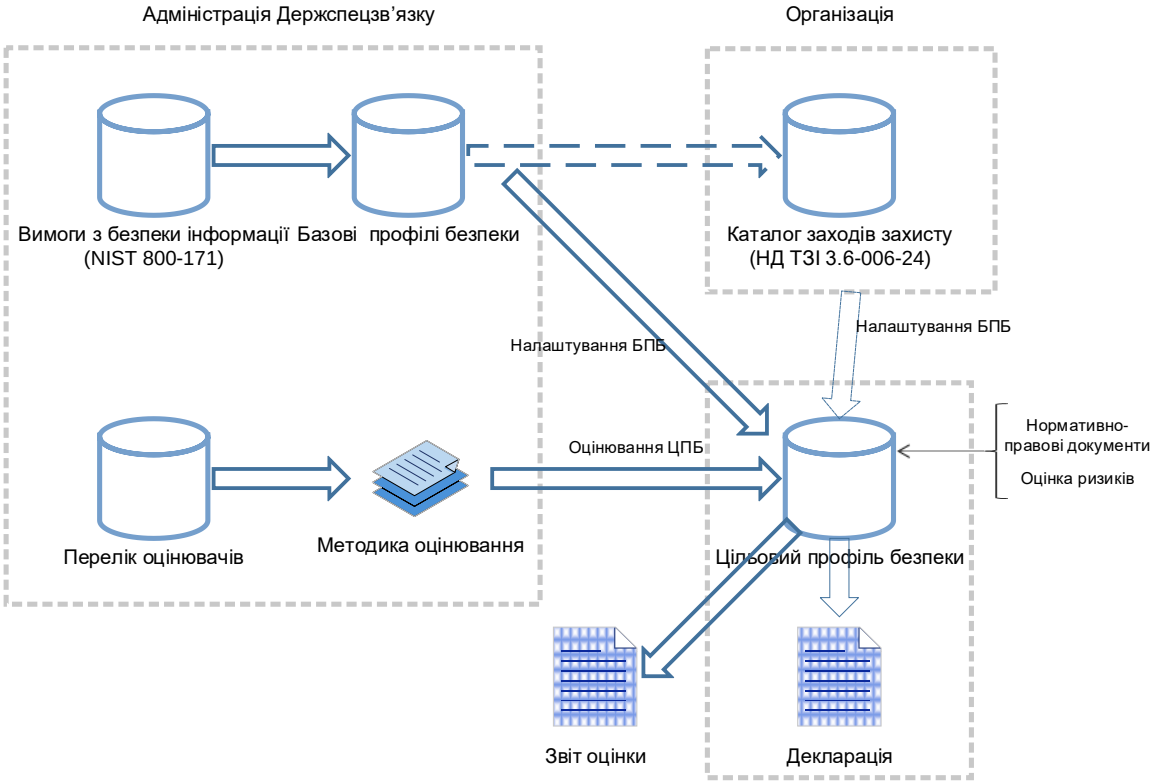


Рис. 2. Застосування профілів безпеки інформації

Передбачені у цільовому профілі заходи із захисту інформації, обрані стандарти, шляхи і способи здійснення таких заходів повинні включати відповідні вимоги та заходи, визначені базовим профілем.

Переваги та недоліки застосування базового профілю безпеки у порівнянні з КСЗІ, СУІБ та СБІ:

Переваги:

1. Простота впровадження: легше впровадити, оскільки він передбачає мінімальні вимоги та контрольні заходи.

2. Низька вартість: зазвичай вимагає менше фінансових та ресурсних витрат порівняно з більш комплексними системами.

3. Швидкість реалізації: може бути швидко розгорнутий, забезпечуючи базовий рівень захисту за короткий час.

4. Зниження ризиків: забезпечує захист від найпоширеніших загроз, що є корисним для організацій з обмеженими ресурсами.

Недоліки:

1. Обмежений рівень захисту: забезпечує лише мінімальний рівень безпеки, що може бути недостатнім для організацій з високими вимогами до безпеки.

2. Необхідність подальшого розширення: для повного захисту інформаційних активів може знадобитися впровадження додаткових заходів безпеки.

3. Менша гнучкість: менше адаптований до специфічних потреб організації порівняно з більш комплексними системами.

3. Процес створення КСЗІ в системах з використанням базових та цільових профілів та декларування відповідності таких комплексних систем

Загальна схема створення КСЗІ в системах з використанням базових та цільових профілів та декларування відповідності таких комплексних систем здійснюється за такими етапами (рис. 3) [4]:

1. Розробка профілів безпеки.
2. Впровадження вимог безпеки.
3. Оцінювання достатності.
4. Декларування відповідності КСЗІ в системах, створених з використанням профілів безпеки.

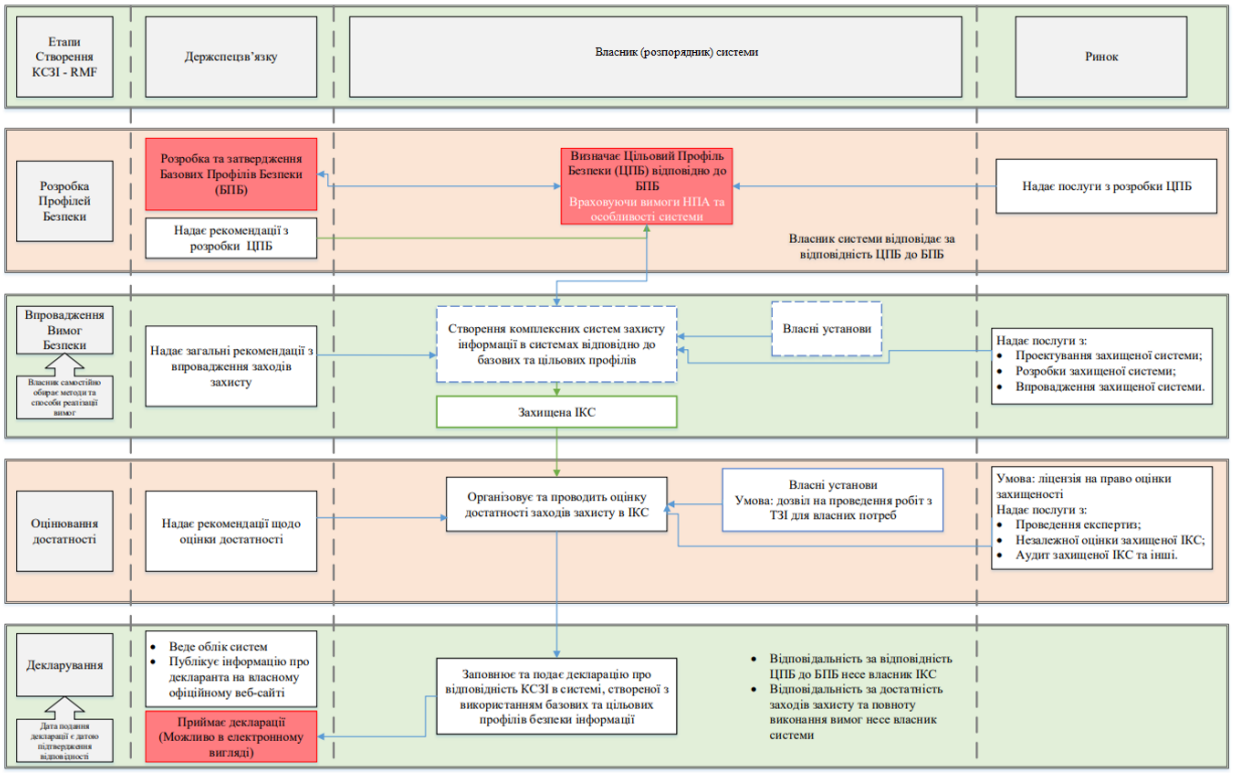


Рис. 3. Порядок реалізації експериментального проєкту

Процес створення КСЗІ в системах з використанням базових та цільових профілів та декларування відповідності таких комплексних систем передбачає виділення трьох ролей: Держспецзв'язку, власника (розпорядника) системи та ринку. В якості останньої ролі мають-ся на увазі державні та приватні підприємства, які надають свої послуги зі створення КСЗІ, наприклад, IT Engineering [11], Державна ІТ-компанія "ІНФОТЕХ" [12], АТ "ІІТ" [13], H-X Technologies [14] та багато інших.

Перший етап – розробка профілів безпеки. На цьому етапі передбачається формування вимог безпеки до системи відповідно до певного виду інформації (відкрита, службова або інформація, що становить державну таємницю) у вигляді базового профілю. Як показано на рис. 1, в першій редакції Адміністрація Держспецзв'язку таку місію виконала у вигляді відповідного наказу від 26.06.2024 № 317 [15]. Важливо також відмітити універсальність, яка закладена в БПБ цього наказу. Важливо також зазначити, що цей БПБ є універсальним та у ньому не буде поділу на АС класу 1, 2, 3. Розроблений БПБ надається власнику системи. У залежності від виду інформації, БПБ містить заходи захисту відповідно до НД ТЗІ 3.6-006-24 [9]. Під час визначення заходів захисту також враховуються профілі

безпеки засобів захисту, вимоги міжнародних та галузевих стандартів. Так, в наказі Адміністрації Держспецзв'язку від 26.06.2024 № 317 [15] надається посилання на 127 заходів захисту (посилених заходів захисту) для відкритої інформації та 153 заходів захисту (посилених заходів захисту) для службової інформації. Таким чином, власник (розпорядник) системи отримує розроблений БПБ. Після цього, власник (розпорядник) системи, враховуючи вимоги нормативно-правових актів певної галузі та особливостей самої ІС, електронної комунікаційної системи та ІКС проводить налаштування ЦПБ на основі БПБ (як показано на рис. 2). Власник (розпорядник) системи для налаштування ЦПБ може звернутися до послуг третьої сторони, але відповідальність покладається виключно на нього. Слід зазначити, що Адміністрація Держспецзв'язку при налаштуванні ЦПБ здійснює лише функції консультаційного характеру та ніякого впливу на власника (розпорядника) системи не проводить. Тому, розроблений на першому етапі ЦПБ затверджується керівником організації, до якої належить система. Саме на керівника покладається юридична відповідальність за правильність налаштування ЦПБ та впровадження заходів захисту інформації певного виду.

Другий етап – впровадження вимог безпеки. Цей етап має суттєве практичне значення для створення КСЗІ на основі налаштованого ЦПБ. Слід зазначити, що наказом визначається мінімально допустимий набір заходів захисту, впровадження яких в КСЗІ є обов'язковим. Але власник (розпорядник) системи самостійно може визначити додаткові вимоги з безпеки та заходи захисту для системи, враховуючи її специфіку. Таким чином, ЦПБ буде містити усі вимоги до певного виду інформації, що прописані в БПБ, а також додаткові вимоги, обумовлені специфікою використання системи. Також на цьому етапі ринок може приймати участь в проєктуванні (розробці) самої захищеної системи та впровадженні вимог безпеки. Знов слід зазначити, що Адміністрація Держспецзв'язку при створенні КСЗІ безпосередньої участі не приймає, лише, за потреби, може надавати консультаційні поради й ніякого впливу на власника (розпорядника) системи не проводить.

Третій етап – оцінювання достатності. Це дуже важливий етап для власника (розпорядника) системи тому, що саме по результатах його проведення він гарантуватиме йому впевненість в реалізації всіх налаштованих (уточнених) заходів захисту, які прописуються в ЦПБ на систему. Згідно з наказом Адміністрації Держспецзв'язку від 10.07.2024 № 354 [16] оцінка достатності заходів захисту інформації комплексних систем захисту інформації включає оцінку:

- визначення цільового профілю безпеки;
- реалізації базового профілю безпеки;
- реалізації цільового профілю безпеки.

Основним результатом на третьому етапі є отримання звіту оцінювання для декларування щодо відповідності реального стану КСЗІ вимогам з безпеки, які визначаються його ЦПБ. Організовує та проводить оцінку достатності власник (розпорядник) системи за допомогою власних фахівців, які мають дозвіл на проведення робіт з ТЗІ або із залученням сертифікованих компаній, які мають ліцензію на право проведення експертизи КСЗІ з оцінкою та аудитом захищеної ІС, ІКС. Знов слід зазначити, що Адміністрація Держспецзв'язку при оцінюванні достатності безпосередньої участі не приймає, лише, за потреби, може надавати консультаційні поради й ніякого впливу на власника (розпорядника) системи не здійснює.

Четвертий (останній) етап – декларування. Це заповнення власником декларації, яка й стане гарантом реалізації БПБ. Згідно з [4], декларація про відповідність КСЗІ, створеної з використанням БПБ та ЦПБ, подається до Адміністрації Держспецзв'язку власником (розпорядником) системи в електронній формі з використанням системи електронної взаємодії органів виконавчої влади з накладенням електронного підпису, що базується на кваліфікованому сертифікаті електронного підпису відповідно до вимог законодавства у сферах електронної ідентифікації та електронних довірчих послуг, а у разі наявності в них інформації з обмеженим доступом подання таких документів здійснюється з дотриманням встановлених правил роботи з документами, які містять інформацію з обмеженим доступом.

Періодичність подання декларацій становить три роки з дати подання декларації до Адміністрації Держспецзв'язку. На Адміністрацію Держспецзв'язку покладається функція ведення обліку задекларованих КСЗІ та публікації інформації про декларанта на своєму веб-сайті, крім матеріалів, що містять інформацію з обмеженим доступом.

Отже, в даному процесі Адміністрація Держспецзв'язку зосереджена лише на етапах розробки профілів безпеки та розробки рекомендацій щодо їх оцінки та надання консультацій (рис. 4).

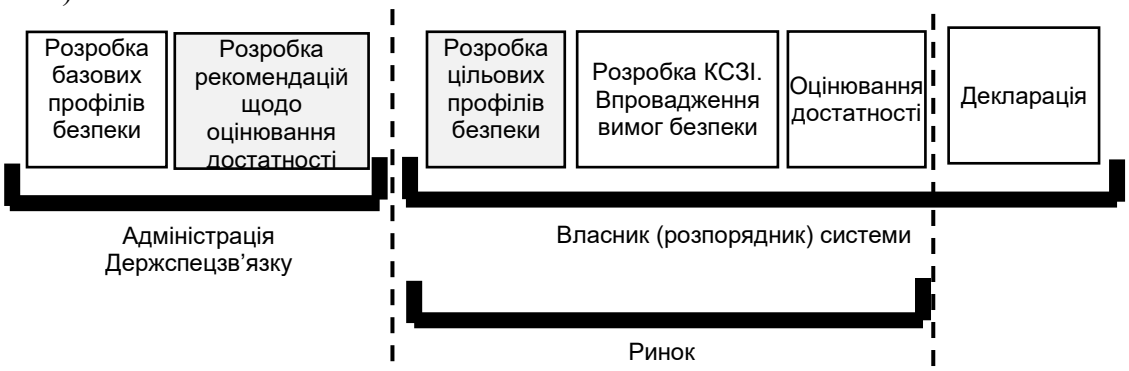


Рис. 4. Етапи, в яких беруть участь учасники експериментального проєкту

Таким чином, після обирання БПБ для відкритої та службової інформації, власник (розпорядник) системи проводить його налаштування. Після затвердження ЦПБ, де він самостійно обирає стандарти, які використовуються під час реалізації заходів захисту, методи та порядок впровадження заходів захисту, починається експлуатація як самої системи, так й КСЗІ, побудованої для неї. Головне правило – набір заходів захисту ЦПБ не може бути менше, ніж встановлено БПБ.

4. Особливі механізми налаштування цільового профілю безпеки

Для налаштування ЦПБ крок за кроком обираються відповідні, для певного виду інформації (відкрита, службова тощо), вимоги з БПБ, які в [15] для відкритої інформації відносяться до 15 класів заходів захисту, а для службової інформації – до 17 класів заходів захисту (табл. 1).

Таблиця 1

Перелік класів заходів захисту, використаних в БПБ для відкритої та службової інформації

№ з/п	ID класу	Назва класу	Кількість вимог в БПБ	
			для відкритої інформації	для службової інформації
1	AC	Управління доступом	16	16
2	AT	Обізнаність і навчання	2	2
3	AU	Аудит і підзвітність	8	8
4	CA	Оцінювання, акредитація та моніторинг безпеки	4	4
5	CM	Управління конфігурацією	6	10
6	IA	Ідентифікація та автентифікація	8	8
7	IR	Реагування на інциденти	4	4
8	MA	Технічне обслуговування	3	3
9	MP	Захист носіїв інформації	6	7
10	PE	Фізичний захист і захист робочого середовища	5	5
11	PL	Планування безпеки	2	3
12	PS	Кадрова безпека	2	2
13	RA	Оцінка ризику	2	2
14	SA	Придбання системи та послуг	-	3
15	SC	Системний і комунікаційний захист	10	10
16	SI	Цілісність системи та інформації	4	5
17	SR	Управління ризиками ланцюга поставок	-	3
		Всього:	82	95

Якщо в якості прикладу розглянути вимогу "1.12 Віддалений доступ" (рис. 5) [9], то для її реалізації власник (розпорядник) системи повинен впровадити заходи захисту, які виділені напівжирним прямокутником (позначені під номером 1). Це заходи захисту АС-17, АС-17(03), АС-17(04) з каталогу НД ТЗІ 3.6-006-24 [9].

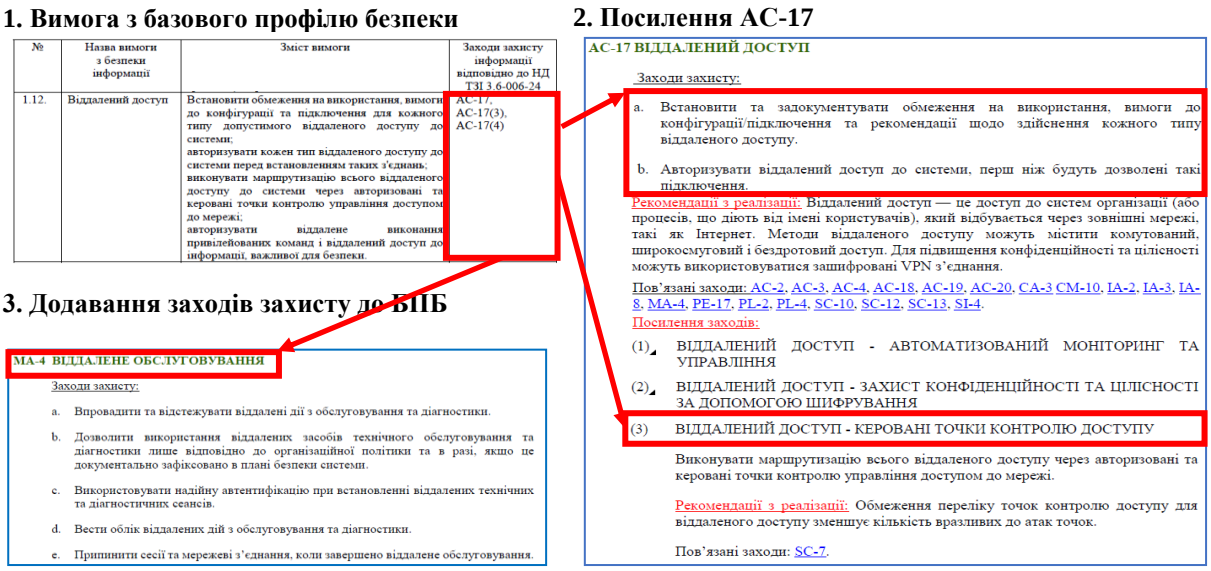


Рис. 5. Формування ЦПБ з посиленням МА-4

Відповідно до нормативно-правових актів організацій і проведення оцінки ризиків умов функціонування системи формується ЦПБ, в якому БПБ буде посилюватися або доповнюватися додатковими заходами захисту. Під номером 2 (рис. 5) зображено захід захисту АС-17 НД ТЗІ 3.6-006-24 [9], який потрібно реалізувати згідно з вимогами БПБ. Напівжирним прямокутником під номером 2 також виділено посилення АС-17(3) заходів захисту, які потрібно виконати та врахувати при налаштуванні ЦПБ. Однак під номером 3 наведений захід захисту МА-4, який не входить до складу БПБ, але може бути використаний при налаштуванні ЦПБ.

Таким чином, аналізуючи всі вимоги (табл. 1) проводиться їх співставлення із заходами захисту, які надаються каталогом НД ТЗІ 3.6-006-24 [9]. Слід також зазначити, що цей каталог є національним гармонізованим стандартом, що відповідає NIST 800-53 rev.5 [3], який періодично змінюється, виходячи з реалій існуючого стану програмних та апаратних можливостей зловмисників.

5. Модель проведення оцінювання достатності

Загальний процес оцінювання достовірності ЦПБ проводиться з метою [16]:

- підтвердження відповідності вибору базового профілю безпеки для формування цільового профілю безпеки;
- підтвердження наявності у цільовому профілі безпеки всіх вимог та заходів захисту базового профілю безпеки;
- підтвердження відповідності цільового профілю безпеки (сукупності заходів із захисту інформації та їх налаштувань) вимогам законодавства та стандартів у сфері захисту інформації, нормативних документів системи технічного захисту інформації, галузевих вимог, політик безпеки тощо для визначеної системи.

У нотації системи умовних позначень для моделювання бізнес-процесів (Business Process Model and Notation, BPMN) [17] процес оцінювання наведено на рис. 6.

Процедура оцінки базового профілю безпеки складається з мети заходу з оцінки та набору потенційних методів і об'єктів оцінки, які можуть бути використані для проведення оцінки. Кожна потенційна мета заходу з оцінки передбачає твердження про визначення, пов'язане з вимогою безпеки, визначеною базовим профілем. Якщо у вимозі з безпеки є параметр, визначений організацією (Organization-Defined Parameter, ODP), то мета заходу

з оцінки починається з формулювання, пов’язаного з визначенням ODP. Визначальні твердження пов’язані із змістом вимог з безпеки, щоб допомогти забезпечити простежуваність результатів оцінки до вимог.

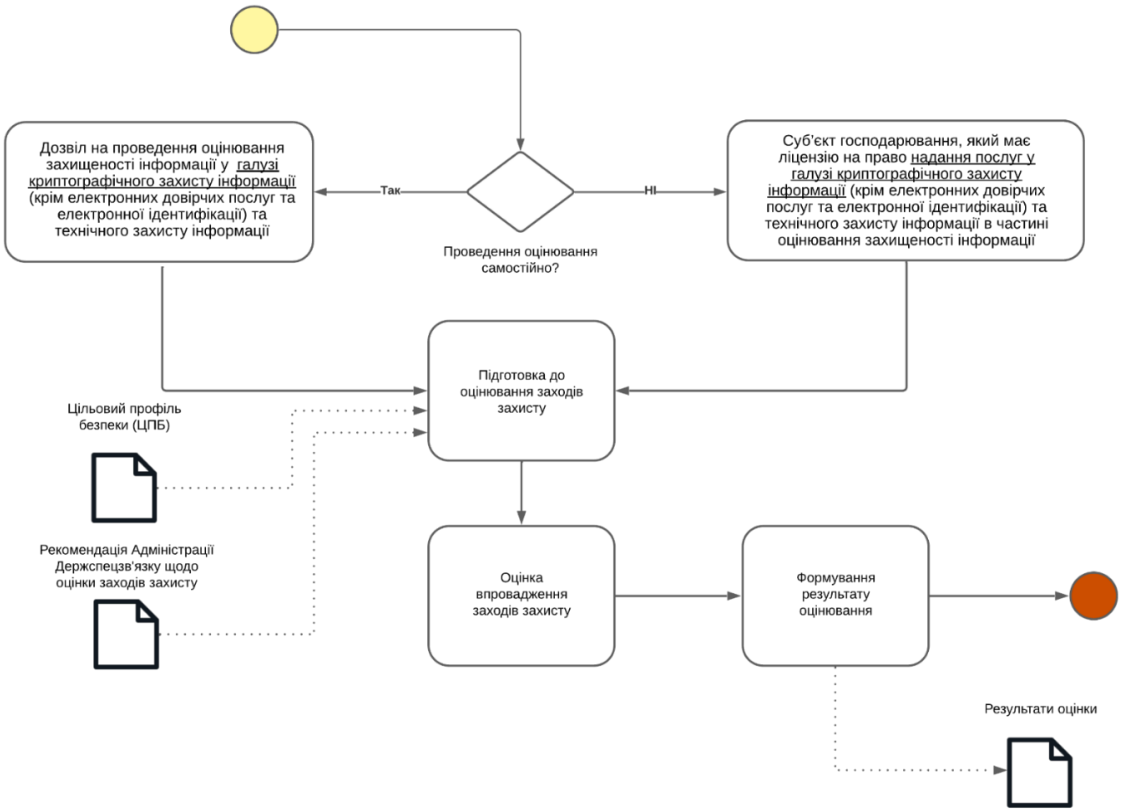


Рис. 6. Модель оцінювання достатності ЦПБ

Методи оцінки визначають характер і обсяг дій оцінювача і використовуються для полегшення розуміння, досягнення роз’яснень або отримання доказів. Потенційні методи оцінки передбачають дослідження, опитування та випробування [16]:

- метод дослідження – процес огляду, вивчення, інспектування, спостереження або аналізу об’єктів оцінки;
- метод опитування – процес проведення дискусій з окремими особами або групами щодо об’єктів оцінки;
- метод випробування – процес виконання об’єктами оцінки (тобто діями, механізмами) певних дій у визначених умовах з метою порівняння фактичної поведінки з очікуваною.

Методи оцінки містять атрибути глибини та охоплення, які визначають ретельність, обсяг і рівень зусиль для проведення оцінки, а також ступінь впевненості в тому, що вимоги з безпеки були виконані.

Розроблена рекомендація щодо оцінювання реалізації БПБ та ЦПБ надається з метою унеможливлення суб’єктивного оцінювання, щоб кожен експерт незалежно один від одного отримував однакові результати. Застосування процедури оцінювання до вимог безпеки дає результати оцінювання або висновки. Ці висновки узагальнюються і використовуються як докази того, що чи має вимога безпеки значення "позитивне" або "негативне".

"Позитивне" (П) вказує на те, що мета оцінювання була досягнута й отримано повністю прийнятний результат.

"Негативне" (Н) – означає, що експерт не зміг отримати достатньо інформації для прийняття рішення, яке вимагається у звіті про результати оцінювання.

У табл. 2 наведено приклад звіту з оцінки вимоги з безпеки 1.8 «Невдалі спроби входу в систему» базового профілю.

Таблиця 2

Фрагмент оцінювання достовірності для вимоги безпеки 1.8 «Невдалі спроби входу в систему»

Номер вимоги	Оцінка		Висновок з оцінки	Докази, джерела отримання відомостей, коментарі оцінювача
	Позначення заходу з оцінки	Мета заходу з оцінки, визначити що:		
1.8	Невдалі спроби входу в систему			
	A.1.8.ODP[01]	визначена кількість послідовних невдалих спроб входу користувача в систему, дозволених протягом певного періоду часу;	П	*Дослідження:* 1. В політиці безпеки організації, яка затверджена наказом директора №111 від 26.06.2024 року, у розділі «Політика використання паролів» визначено кількість послідовних недійсних спроб входу до системи – 5 спроб; 2. В документі «Технічний проєкт. Пояснювальна записка» (реєстр. № 123 від 26.06.2024 року) у відповідності до пункту 4 розділу 3 політика використання паролів реалізується засобами ОС Windows 10 (параметр налаштування «Account lockout threshold»). 3. У налаштуваннях ОС Windows 10 системи параметр «Account lockout threshold» встановлено «5». *Опитування:* 1. Адміністратор безпеки – здійснив налаштування системи у відповідності до «Інструкції адміністратора безпеки» (реєстр. № 121 від 26.06.2024 року) та документу «Технічний проєкт. Пояснювальна записка» (реєстр. № 123 від 26.06.2024 року). *Випробування:* 1. Після 5 спроб введення неправильного паролю до ОС Windows 10 система блокується.
	A.1.8.ODP[02]	визначено період часу, яким обмежено кількість послідовних невдалих спроб входу користувача;	Н	*Дослідження:* 1. У політиці безпеки організації, яка затверджена наказом директора №111 від 26.06.2024 року, у розділі «Політика використання паролів» визначено період часу, яким обмежено кількість послідовних невдалих спроб входу користувача – 5 хвилин. 2. Механізми реалізації вимоги **не визначені.** 3. Спосіб налаштування механізмів **не визначено.** *Опитування:* 1. Адміністратор безпеки – відповідні налаштування системи **не проводив**. *Випробування:* 1. В продовж 7 хвилин здійснено 4 спроби введення неправильного паролю до ОС Windows 10 системи – система **не блокується.**
	A.1.8	визначено кількість послідовних невірних спроб входу користувача протягом <A.1.8.ODP[02]: період часу> обмежено до <A.1.8.ODP[01]: кількість>.	Н	*Коментарі:* 1. Вимога не може бути оцінена – A.1.8.ODP[02] **оцінена негативно**.
	Загальна оцінка реалізації вимоги 1.8		**Не реалізовано**	

У табл. 2 «Номер вимоги» відповідає номеру вимоги з безпеки базового профілю для систем, де обробляється відповідна інформації за видом доступу.

Позначення заходів з оцінки мають буквено-цифрові ідентифікатори. Кожний захід з оцінки починається з літери «А», яка вказує на те, що воно є частиною процедури оцінки. Наступна послідовність цифр та/або літер (наприклад, 1.8.ODP[01]) вказує на ідентифікатор вимоги безпеки з базового профілю для ІКС, де обробляється відповідна інформації за видом доступу (та конкретний елемент контролю, якщо це багатокомпонентна вимога), яка є метою заходу з оцінки. Параметри, визначені організацією, позначаються літерами «ODP». Якщо в заяві про визначення є декілька ODP, номер ODP вказується в квадратних дужках (наприклад, A.1.8.ODP[01]). Квадратні дужки також використовуються для позначення того, коли процедура оцінки далі розбиває вимогу на більш детальні заяви про визначення (наприклад, A.1.8.ODP[01], A.1.8.ODP[02]).

Порядок оцінки КСЗІ на основі ЦПБ визначається власником системи самостійно та здійснюється суб'єктами господарювання, які мають ліцензію на право надання послуг в галузі криптографічного захисту інформації (крім надання електронних довірчих послуг та електронної ідентифікації) та технічного захисту інформації в частині оцінювання захищеності інформації або учасниками експериментального проєкту, які мають дозвіл на проведення робіт з ТЗІ для власників в частині захищеності інформації.

6. Модель декларування відповідності

Результат оцінки КСЗІ на основі ЦПБ, що підготовлено в рамках цього експериментального проєкту, приймається як результат державної експертизи. За умов виконання всіх етапів та вимог, власник системи декларує, що набір заходів захисту, визначених ЦПБ, відповідає мінімальному набору заходів захисту, встановленому в БПБ для відповідного виду інформації. В нотації BPMN процес декларації показаний на рис. 7.

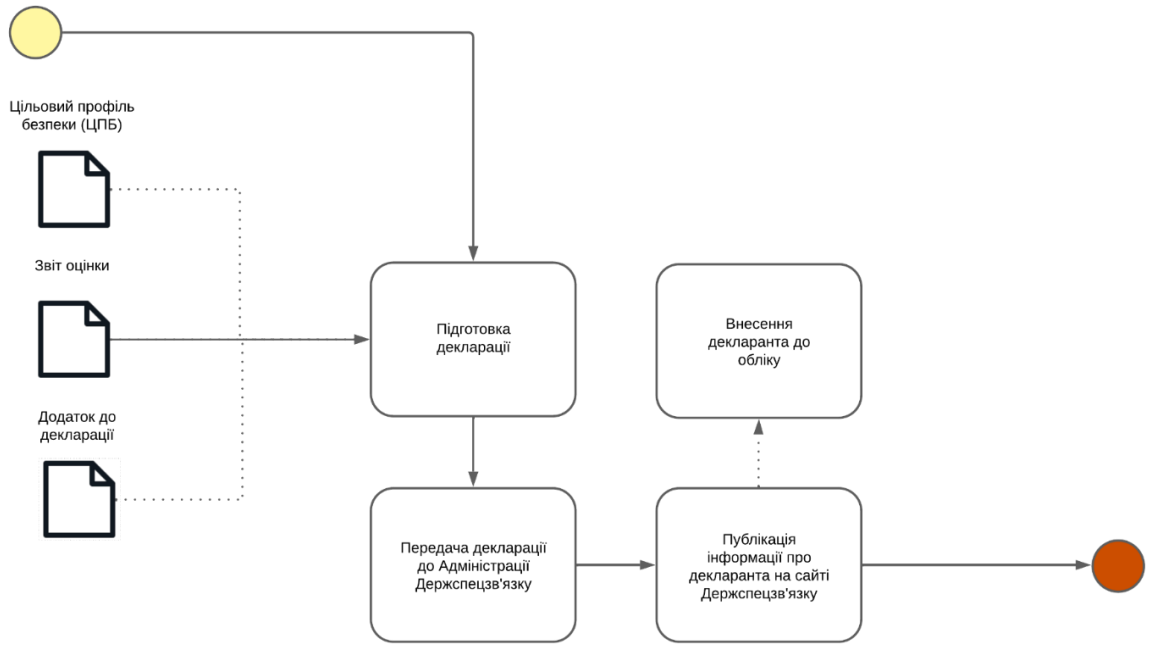


Рис. 7. Модель декларування відповідності

Декларація про відповідність КСЗІ подається до Адміністрації Держспецзв'язку власником системи в електронній формі з накладанням кваліфікованого електронного підпису, а для документів з обмеженим доступом – в установленому законодавством порядку. Облік систем, які мають КСЗІ, побудованих з використанням профілів безпеки, буде здійснюватися Адміністрацією Держспецзв'язку.

Дата на відсилання декларації про відповідність до Адміністрації Держспецзв'язку буде вважатися датою підтвердження відповідності цієї КСЗІ в системах, побудованих з використанням профілів безпеки та не потребує відповіді від Адміністрації Держспецзв'язку.

Адміністрація Держспецзв'язку буде публікувати інформацію про декларантів на сайті Держспецзв'язку. Щодо перевірки КСЗІ в системах, побудованих з використанням профілів безпеки, то вона буде здійснюватися в рамках державного контролю за станом технічного захисту інформації державних інформаційних ресурсів та інформації, вимоги до захисту яких встановлені законом.

Адміністрація Держспецзв'язку, з метою визначення ефективності реалізації експериментального проєкту, здійснює моніторинг системи захисту інформації в системах, побудованих з використанням даних профілів. Моніторинг проводиться з метою надання методичної консультативної підтримки. Строк дії декларації становить три роки з дня підтвердження її відповідності [4].

У разі закінчення терміну дії декларації про відповідність або зміну ЦПБ, декларація про відповідність КСЗІ вважатиметься недійсною. Додаткове декларування КСЗІ в системах, побудованих з використанням профілів безпеки, відбувається за наступних підстав: закінчення терміну дії декларації про відповідність, або в разі зміни ЦПБ, або після усунення недоліків, виявлених за результатами проведення державного контролю.

Висновки

1. Розглянутий підхід декларування дозволить поєднати сучасні нормативно-правові документи, які утворюють систему використання профілів безпеки інформації при розгортанні КСЗІ, та спрощення отримання підстав експлуатувати інформаційні, електронні комунікаційні та інформаційно-комунікаційні системи з обробкою відкритої та службової інформації.

2. Розглянутий підхід дозволить підвищити рівень обізнаності проєктувальників та розробників КСЗІ, надавши підготовлений заздалегідь базовий профіль безпеки та вказавши шлях його застосування.

3. Побудовано логічну сукупність нормативно-правових актів, яка дозволить орієнтуватися в сучасних підходах, які ґрунтуються на стандартах NIST США.

4. Основна увага у роботі приділена процесу розробки профілів безпеки, впровадженню вимог безпеки, механізмам оцінювання достатності та подальшого декларування відповідності КСЗІ в системах, створених з використанням профілів безпеки.

Список літератури:

1. ISO/IEC 27000 family. Information security management. URL: https://www.iso.org/standard/iso-iec-27000-family
2. НД ТЗІ 3.6-004-21. Порядок впровадження системи безпеки інформації в державних органах, на підприємствах, організаціях, в інформаційно-комунікаційних системах яких обробляється інформація, вимога щодо захисту якої встановлена законом та не становить державної таємниці URL: https://cip.gov.ua/services/cm/api/attachment/download?id=53375
3. NIST SP 800-53 rev. 5 Security and Privacy Controls for Information Systems and Organizations, 2020. URL: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r5.pdf
4. Постанова Кабінету Міністрів України від 30.05.2024 № 627 "Про реалізацію експериментального проєкту з декларування відповідності комплексних систем захисту інформації в інформаційних, електронних комунікаційних та інформаційно-комунікаційних системах, створених з використанням профілів безпеки інформації". URL: https://zakon.rada.gov.ua/laws/show/627-2024-п.
5. НД 1.4-001-2000. Типове положення про службу захисту інформації в автоматизованій системі, 2000. URL: https://tzi.com.ua/downloads/1.4-001-2000.pdf
6. NIST SP 800-53A rev. 5 Assessing Security and Privacy Controls in Information Systems and Organizations, 2022. URL: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53Ar5.pdf
7. NIST SP 800-171 rev. 3. Protecting Controlled Unclassified Information in Nonfederal Systems and Organizations, 2024. URL: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-171r3.pdf
8. NIST SP 800-171A rev. 3. Assessing Security Requirements for Controlled Unclassified Information, 2024. URL: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-171Ar3.pdf.

9. НД ТЗІ 3.6-006-24. Порядок вибору заходів захисту інформації, вимога щодо захисту якої встановлена законом та не становить державної таємниці, для інформаційних систем, 2024. URL: https://cip.gov.ua/services/cm/api/attachment/download?id=64620
10. НД ТЗІ 2.3-025-21. Методика оцінювання заходів захисту інформації, вимога щодо захисту якої встановлена законом та не становить державної таємниці, для інформаційних систем, 2021. URL: https://cip.gov.ua/services/cm/api/attachment/download?id=53377
11. IT Engineering. URL: https://it-engineering.com.ua
12. Державна IT-компанія "ІНФОТЕХ" URL: https://infotech.gov.ua.
13. АТ "ІІТ" URL: https://infotech.gov.ua
14. H-X Technologies URL: https://www.h-x.technology.
15. Наказ Адміністрації Держспецзв'язку від 26.06.2024 "Про визначення Базового профілю безпеки інформації". URL: https://www.cip.gov.ua/ua/news/nakaz-administraciyi-derzhspeczv-yazku-vid-24-06-2024-317-pro-viznachennya-bazovogo-profilyu-bezpeki-informaciyi
16. Наказ Адміністрації Держспецзв'язку від 10.07.2024 "Про затвердження Рекомендацій з оцінки достатності заходів захисту інформації комплексних систем захисту інформації в інформаційних, електронних комунікаційних та інформаційно-комунікаційних системах, створених з використанням профілів безпеки інформації". URL: https://cip.gov.ua/ua/news/nakaz-administraciyi-derzhspeczv-yazku-vid-10-07-2024-354-pro-zatverdzhennya-rekomendacii-z-ocinki-dostatnosti-zakhodiv-zakhistu-informaciyi-kompleksnikh-sistem-zakhistu-informaciyi-v-informaciinikh-elektronnikh-komunikaciinikh-ta-informaciino-komunikaciinikh-sistemakh-stvorenikh-z-vikoristannyam-profiliv-bezpeki-informaciyi
17. Система умовних позначень (нотація) для моделювання бізнес-процесів версії 2.0. URL: https://uk.wikipedia.org/wiki/BPMN

Надійшла до редколегії 10.06.2024

Відомості про авторів:

Потій Олександр Володимирович – д-р техн. наук, професор, заступник Голови Державної служби спеціального зв'язку та захисту інформації України; e-mail: Potav1971@gmail.com4 ORCID: https://orcid.org/0009-0004-9332-4414

Голубничий Дмитро Юрійович – канд. техн. наук, доцент, Харківський національний економічний університет імені Семена Кузнеця, доцент кафедри інформаційних систем, факультет інформаційних технологій, АТ "Інститут Інформаційних Технологій", начальник відділу наукових досліджень; Україна; e-mail: dmytro.holubnychyi@hneu.net, ORCID: https://orcid.org/0000-0002-6873-7004

Васильєв Юрій Костянтинович – Державна служба спеціального зв'язку та захисту інформації України, e-mail: y.vasyliev@cip.gov.ua

Єсіна Марина Віталіївна – канд. техн. наук, доцент, Харківський національний університет імені В. Н. Каразіна, доцент кафедри безпеки інформаційних систем і технологій, факультет комп'ютерних наук, АТ "Інститут Інформаційних Технологій", науковий співробітник-консультант; Україна; e-mail: m.v.yesina@karazin.ua; ORCID: https://orcid.org/0000-0002-1252-7606

УДК 004.056.5 **DOI:10.30837/rt.2024.2.217.02**

Я.А. ДЕРЕВ'ЯНКО, М.В. ЄСІНА, канд. техн. наук, Д.Ю. ГОРБЕНКО

ОБҐРУНТУВАННЯ МЕТОДІВ ОБЧИСЛЕННЯ ТА АНАЛІЗ ВЛАСТИВОСТЕЙ ПСЕВДОВИПАДКОВИХ ТА ВИПАДКОВИХ ПОСЛІДОВНОСТЕЙ НА ОСНОВІ ДНК

Вступ

Невід'ємною вимогою до сучасних інформаційних систем є надання користувачам послуг конфіденційності, цілісності, доступності та неспростовності. Якість таких послуг напряму залежить від криптографічних перетворень, важливою складовою яких є випадковість. Тому генерування псевдовипадкових (ПВП) та випадкових (ВП) послідовностей є однією з актуальних та важливих задач. Випадкові числа використовуються для генерації сеансових ключів, параметрів підпису, nonce, викликів, засліплення та маскування значень (для запобігання атакам на реалізацію) тощо. ВП генеруються на основі фізичних та нефізичних джерел шуму [1]. ПВП генеруються з використанням генераторів випадкових послідовностей (ГВП), як правило на основі відносно коротких ВП, наприклад, ключів тощо.

Наші попередні дослідження [2] вказують на теоретичну можливість використання у якості джерела шуму (ДШ) та відповідно джерела ВП ДНК. Як показав пошук та аналіз, детальних досліджень в цьому напрямі немає чи вони недоступні. Також згідно [1] ДНК можливо віднести до нефізичних ДШ та відповідно ГВП. Для оцінки та порівняння ПВП та ВП на основі використання ДНК потрібно провести широкі практичні та теоретичні дослідження з використанням ентропійних (стохастичних) методів та методик, та визнаних статистичних методик (бажано стандартизованих). Цим вимогам задовольняють AIS 20 та AIS 31 [1], що визначають стандартизовані удосконалені методи стохастичного та статистичного оцінювання, та порівняння з іншими ДШ та відповідно ГВП, як потрібно оцінювати ГВЧ.

Дана стаття присвячена новим розробленим методам отримання ПВП та ВП на основі послідовностей ДНК. Вони розглядаються у якості не фізичних справжніх ВП, в тому числі з використанням за необхідності екстракторів [1, 2]. Результатами дослідження є згенеровані ПВП та ВП на основі ДНК, а також експериментально отримані значення статистичної оцінки та оцінки подібності послідовностей.

По суті ця стаття є вступом в теорію та практику генерування ПВП та ВП на основі ДНК. У ній подаються результати вирішення наступних проблемних питань:

1) Пропозиції щодо інтерпретації та подання ДНК у якості не фізичного ДШ та подальшої оцінки.

2) Обґрунтування та розробка методів обчислення ПВП та ВП, а також методів порівняння послідовностей.

3) Аналіз статистичних та стохастичних властивостей ВП та ПВП на основі ДНК, а також удосконалення їх властивостей на основі екстракції.

4) Аналіз подібності ПВП, ВП та ДНК послідовностей для різних ДНК.

Вважаємо, що актуальними та необхідними є подальші дослідження, оцінка та порівняння різних ДНК, встановлення подібності ПВП різних ДНК тощо.

1. Інтерпретація ДНК для можливості виконання подальшого дослідження та оцінка властивостей таких послідовностей

У ДНК зустрічається чотири види азотистих основ (аденін, гуанін, тимін і цитозин). Зважаючи на це ДНК можна представити у вигляді послідовності азотистих основ наступним чином (рис. 1):

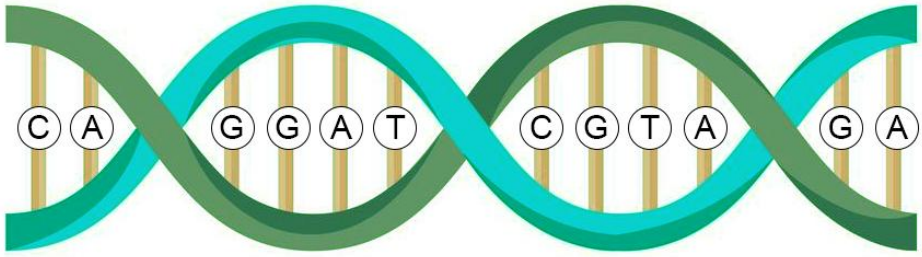


Рис. 1. ДНК (coding strand) у вигляді азотистих основ

Таким чином, маємо послідовність, яка складається з алфавіту, у якому можливі чотири значення – A, C, G та T. Така послідовність може бути представлена у вигляді двійкової послідовності шляхом заміни цих значень на відповідні двійкові комбінації – A=00, C=01, G=10 та T=11. Наприклад, маючи таку послідовність як на рис. 1 – CAGGATCGTAGA – отримаємо двійкову послідовність 010010100011011011001000.

Зразки ДНК різних організмів різної довжини доступні на сайті Національної Бібліотеки Медицини США у банку генів (GenBank) [3] та у ДНК банку Японії (DNA Data Bank of Japan) [4].

Для проведення досліджень було обрано декілька зразків ДНК різних організмів. Обрані послідовності ДНК було представлено у вигляді двійкових як описано вище. Для кожної з них було проведено статистичне та стохастичне тестування. Отримані результати надають інформацію про те чи мають «сирі» послідовності достатньо випадковості і чи потребують вони покращення.

Результати тестування «сирих» послідовностей представлено на прикладі тестування послідовностей з ДНК людини, оскільки результати для інших послідовностей є доволі схожими. Оскільки показники не є задовільними, їх представлення у вигляді статистичного портрету не є доцільним (рис. 2, 3):

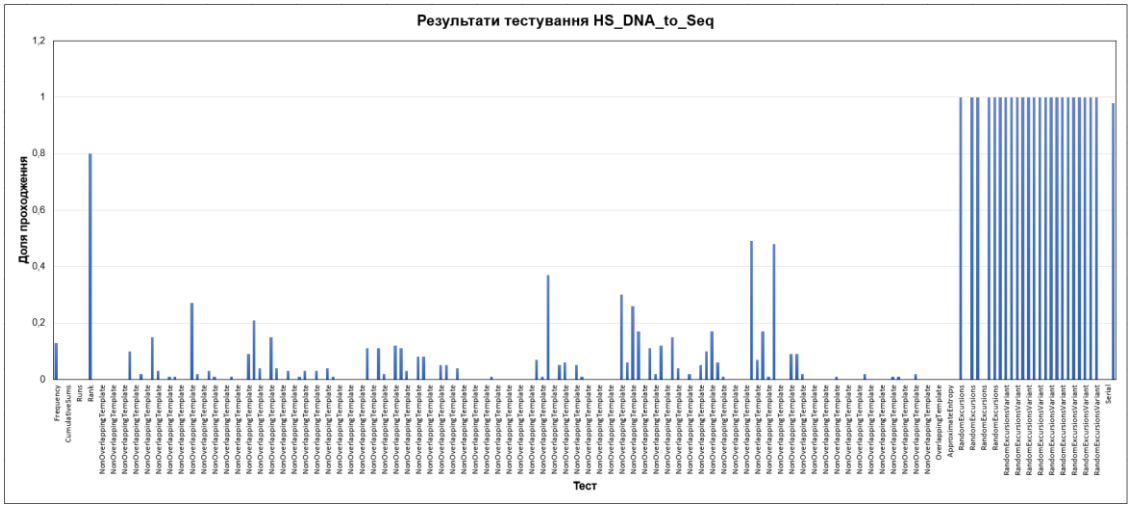


Рис. 2. Результати статистичного тестування «сирої» послідовності з ДНК людини

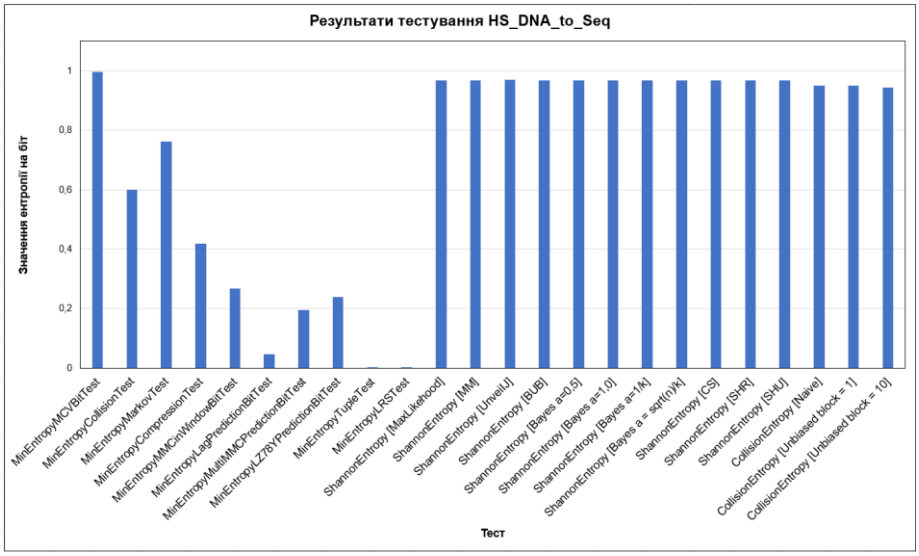


Рис. 3. Результати стохастичного тестування «сирої» послідовності з ДНК людини

Як видно з отриманих результатів, статистичні характеристики «сирих» послідовностей не є задовільними, майже всі статистичні тести провалені. Хоча більшість тестів мінімальної ентропії мають незадовільне значення, показники ентропії Шеннона та колізійної ентропії є хорошими, що вказує на присутність певної кількості ентропії у таких послідовностях, це дає можливість застосування екстракції випадковості з використанням, наприклад, блокових симетричних шифрів (БСШ).

2. Обґрунтування методів обчислення ПВП та ВП, а також методів порівняння послідовностей

Для отримання послідовностей було використано екстрактор або ж DRNG на основі ДСТУ 7624:2014 [5] з довжиною ключа 256 біт у режимі гамування (CTR), оскільки саме такий режим пропонується використовувати з блоковим шифром для отримання ПВП та ВП у AIS 20/31 [1].

У стандарті [5] передбачено наступне:

Для генерації псевдовипадкових послідовностей з використанням ДСТУ 7624:2014 у режимі гамування (CTR):

- Задання ключа шифрування
  для λ=256 Key = Hash256 (b || 1, bBytes+1)
  для λ>256 Key = Hash512 (b || 1, bBytes+1)
- Задання синхропосилки
  для λ=256 IV = Hash256 (b || 2, bBytes+1)
  для λ>256 IV = Hash512 (b || 2, bBytes+1),

де b – рядок октетів, який визначає внутрішній стан генератора, bBytes (октетів) – його довжина.

Для отримання випадкових послідовностей, необхідно застосовувати у якості ключа та синхропосилки випадкові значення. У стандарті [5] вказано наступне.

Для генерації випадкових послідовностей з використанням ДСТУ 7624:2014 в режимі гамування (CTR):

- як ключ шифрування вибираємо випадковий двійковий рядок завдовжки 256, якщо λ=256 та 512, якщо λ=384 або λ = 512. Цей ключ визначає внутрішній стан генератора;
- як синхропосилку вибираємо випадковий двійковий рядок завдовжки 256, якщо λ=256 та 512, якщо λ=384 або λ=512.

Таким чином, ДСТУ 7624:2014 дозволяє отримати ПВП та ВП в залежності від обраного режиму. Далі будуть представлені алгоритми отримання ПВП та ВП, в яких враховано виконання вказаних вище вимог.

2.1. Отримання псевдовипадкових послідовностей з ДНК з використанням ДСТУ 7624:2014 на одному ключі

Для отримання ПВП було використано екстрактор на основі ДСТУ 7624:2014 з довжиною ключа 256 біт у режимі гамування (CTR).

Псевдокод алгоритму отримання псевдовипадкових послідовностей наводиться далі:

```
Алгоритм 1
Вхід:
Файл з необробленими даними
Вихід:
Файл з псевдовипадковою послідовністю на виході
1 Вибір режиму роботи
#define CK_SIZE 32
2 Ініціалізація контексту
kalyna_t* ctx44_e = KalynaInit(256, 256);
3 Ініціалізація буферу, в який буде зчитуватися необроблена послідовність
uint64_t mem[seq_len/8];
4 Зчитування послідовності з бінарного файлу у буфер
mem[i] = байти[i – i+8]
5 Ініціалізація ключа та синхропосилки залежно від режиму
uint64_t initKey …._e[CK_SIZE/8] = {0x.., 0x.., … 0x… };
uint64_t nonce …._e[CK_SIZE/8] = {0x.., 0x.., … 0x… };
6 Набір послідовності необхідної довжини
while (не досягнуто необхідної довжини послідовності)
Зашифрування початкової/вже зашифрованої раніше послідовності
KalynaKeyExpand(key, ctx)
counter = 0;
        for (i < dataLen; i+= CK_SIZE/8){
                uint64_t pt…._e[CK_SIZE/8] = {nonce[i] … nonce[i+…]};
                pt = pt^counter;
                counter++;
                KalynaEncipher(pt…._e, ctx…._e, ct…._e);
                ct … ct[i+…]  = ct^mem;
                mem[i] … mem[i+…] = ct…_e[i] … ct…_e[i+…];
        }
        Запис у файл зашифрованої частини
end while
```

Розроблений алгоритм дозволяє отримувати ПВП на основі вхідних послідовностей та обраних ключа і синхропосилки. Таким методом було отримано послідовності різних довжин з ДНК різних організмів. Результати аналізу статистичних властивостей таких послідовностей будуть надані у наступному пункті.

2.2. Отримання випадкових послідовностей з ДНК з використанням ДСТУ 7624:2014 на сеансових ключах, отриманих з NPTRNG ядра Linux

Для відповідності вимогам [5] щодо ключа та синхропосилки для генерації ВП, їх отримання відбувалося з використанням NPTRNG /dev/random [6] на ядрі Linux 5.4, що за AIS 20/31 [1] відповідає класу функціональності NTG.1, проте в більш пізніх версіях ядра 5.5+ цей генератор відносять до DRNG класу DRG.3 [1].

Отже, отримання випадкової послідовності можливе подібним чином:

1. Обираємо з ДНК людини двійкову послідовність невеликого розміру (з метою частішої зміни ключа шифрування та синхропосилки), наприклад 1 МБ.

2. З використанням NPTRNG отримуємо ключ шифрування та синхропосилку для кожної ітерації зашифрування (наприклад, з використанням генератора /dev/random на платформі Linux).

3. Зашифровуємо послідовність ДНК необхідну кількість разів, щоб отримати необхідну довжину вихідної послідовності.

Псевдокод алгоритму отримання випадкових послідовностей наводиться далі.

```
Алгоритм 2
Вхід:
Файл з необробленими даними
Вихід:
Файл з випадковою послідовністю на виході
1 Вибір режиму роботи
#define CK_SIZE 32
2 Ініціалізація контексту
kalyna_t* ctx44_e = KalynaInit(256, 256);
3 Ініціалізація буферу, в який буде зчитуватися послідовність
uint64_t mem[seq_len/8];
4 Зчитування послідовності з бінарного файлу у буфер
mem[i] = байти[i – i+8]
5 Набір послідовності необхідної довжини
while (не досягнуто необхідної довжини послідовності)
        Ініціалізація ключа та синхропосилки залежно від режиму
        fd = open("/dev/random", O_RDONLY);
        read(fd, initKey, CK_SIZE);
        read(fd, nonce, CK_SIZE);
        Зашифрування початкової/вже зашифрованої раніше послідовності
        KalynaKeyExpand(key, ctx)
        counter = 0;
        for (i < dataLen; i+= CK_SIZE/8){
                uint64_t pt…._e[CK_SIZE/8] = {nonce[i] … nonce[i+…]};
                pt = pt^counter;
                counter++;
                KalynaEncipher(pt…._e, ctx…._e, ct…._e);
                ct[i] … ct[i+…]  = ct^mem;
                mem[i] … mem[i+…] = ct…_e[i] … ct…_e[i+…];
        }
        Запис у файл зашифрованої частини
end while
```

Розроблений алгоритм дозволяє отримувати ВП на основі вхідних послідовностей та ключа і синхропосилки, отриманих в використанням відповідного генератора. Таким методом також було отримано послідовності різних довжин з ДНК різних організмів. Результати статистичного аналіз представлено у наступному пункті.

Далі розглянуто методи порівняння послідовностей.

2.3. Порівняння послідовностей з використанням підрахунку k-мерів та k-мер відстані

Хорошим методом порівняння без попереднього вирівнювання ДНК послідовностей або будь-яких рядків на подібність є підрахунок k-мерів. У біоінформатиці k-мери – це підрядки довжиною *k*, що містяться в біологічній послідовності. Переважно використовуються в контексті обчислювальної геноміки та аналізу послідовностей.

При застосуванні, наприклад, добітових чи байтових потоків, або простих рядків у якості k-мерів виступають підрядки довжини *k*: для бітового рядка підрядки бітів, для байтового – байтів, а для звичайного тесту – підрядки символів відповідно.

У якості прикладу для демонстрації взято частину послідовності ДНК довжиною 8 нуклеотидів. У табл. 1 показано всі можливі k-мери довжини від 1 до 8 для заданої послідовності.

Таблиця 1

Значення всіх можливих k-мерів (підрядків довжиною *k*)
для демонстраційної послідовності

CACGATCG	
Значення k	k-мери
1	C, A, C, G, A, T, C, G
2	CA, AC, CG, GA, AT, TC, CG
3	CAC, ACG, CGA, GAT, ATC, TCG
4	CACG, ACGA, CGAT, GATC, ATCG
5	CACGA, ACGAT, CGATC, GATCG
6	CACGAT, ACGATC, CGATCG
7	CACGATC, ACGATCG
8	CACGATCG

Розкладання послідовності на k-мери для аналізу дозволяє аналізувати набір фрагментів фіксованого розміру, а не саму послідовність, що може бути більш ефективним. k-мери дуже корисні для зіставлення послідовностей, а операції з множинами швидші, простіші, і для роботи з ними існує багато доступних алгоритмів і технік. По суті, використання k-мерів спрощує біоінформатику до підрахунку та порівняння наявності чи відсутності речей.

Отже, для порівняння двох ДНК послідовностей для кожної з них спочатку підраховується кількість k-мерів у її складі. Створюється таблиця, де кожному з k-мерів (4^k для стандартного алфавіту ДНК і, наприклад, 2^k у випадку двійкового алфавіту) відповідає кількість входжень конкретного k-меру у послідовність, що досліджується.

Для такого підрахунку можна також скористатися пакетом kmer для мови програмування R [7], який підраховує кількість k-мерів у послідовностях, проте у пакеті підраховуються всі можливі k-мери простору 4^k. Такий метод спричиняє серйозне навантаження на апаратні ресурси комп'ютера, найбільшим чином на оперативну пам'ять при великих значеннях *k*. Оскільки для збереження масиву k-мерів для k=31 потрібно 4.6^{18} елементів, що не є можливим на жодному з сучасних комп'ютерів.

У даному дослідженні пропонується новий метод, який є доволі швидким, не потребує великих затрат пам'яті та має порівняно невисоку алгоритмічну складність.

Запропонований метод полягає в наступному.

Алгоритм 3

1 Пошук всіх можливих k-мерів у обох послідовностях

2 Створення об'єднання цих k-мерів, щоб вилучити зі списку повторні k-мери

3 Представлення k-мерів у двійковому вигляді (A=00, C=01, G=10, T=11). Це в свою чергу дає можливість представлення отриманих k-мерів у вигляді 64-бітних чисел, якщо k<33 для ДНК послідовностей і k<65 для двійкових послідовностей

4 Створення map для кожної з послідовностей

5 Присвоєння в якості ключа значень з набору різних (distinct) k-мерів

6 Прохід по послідовностях і додавання +1 до індексу з ключем, який відповідає k-меру, що знайдено в послідовності. Такий прохід не потребує подвійного циклу і дозволяє підраховувати кількість входжень кожного з k-мерів під час одного проходження

7 Розрахунок k-мер відстані

Математичну формулу для розрахунку k-мер відстані було представлено у роботі [8]. При реалізації алгоритму було використано формулу, що використовується для розрахунку k-мер відстані у пакеті kmer для мови програмування R [7], а саме:

$$F = \sum_{Distinct.length} \frac{\min(p(s_1), p(s_2))}{\min(len(s_1), len(s_2)) - k + 1}, \quad dist(s_1, s_2) = \frac{\log(0.1 + F) - \log(1.1)}{\log(0.1)} \tag{1}$$

де F – дробове загальне число k-мер, $p(s)$ – відповідний k-мер з простору 4^k (2^k) кожної з послідовностей, а $len(s)$ – довжина послідовностей.

Для коректного вибору розміру k-мерів необхідно також враховувати розмір послідовностей, що будуть порівнюватися, а також алфавіт цих послідовностей. Інструмент Mash [9] пропонує визначати розмір k-мерів для оцінки, оцінюючи ймовірність випадкового збігу як

$$p = \frac{1}{\frac{\left(\Sigma\right)^k}{g} + 1}, \tag{2}$$

де g – розмір геному (послідовності), а Σ – алфавіт (ACGT або, наприклад, 01). Якщо ця ймовірність перевищує поріг (за замовчуванням 0.01), то розмір k-мерів – k – необхідно збільшувати.

У інших джерелах, пропонується обирати розмір так, щоб загальна кількість доступних k-мерів була достатньо більшою за розмір геному, що досліджується.

2.4. Порівняння послідовностей на основі MinHash відстані з використанням геш-функції Купина (ДСТУ 7564:2014)

Ще одним методом, який забезпечує досить високу швидкість обчислення, а також має невисокі вимоги до пам'яті, що буде використовуватися, є алгоритм MinHash. У комп'ютерних науках та аналізі даних MinHash – це техніка для швидкої оцінки того, наскільки схожі два набори. Схема була вперше запропонована Andrei Broder у роботі [10]. Застосування такого методу до порівняння послідовностей ДНК вперше згадується у роботі [11]. У нашому дослідженні пропонується порівнювати як послідовності ДНК, так і двійкові послідовності на основі алгоритму MinHash з використанням у якості геш-функції національного стандарту ДСТУ 7564:2014 [12].

Метод порівняння двох послідовностей передбачає наступний порядок дій.

Алгоритм 4

1 Розбиття послідовностей на k-мери необхідної довжини

У нашому дослідженні пропонується перетворювати k-мери на бітові послідовності, що дасть можливість зберігати їх у вигляді 64-бітних чисел. Це, в свою чергу, дозволить розбивати послідовності ДНК на k-мери довжини до 32 включно, що є достатнім для будь-якого геному, а двійкові послідовності на k-мери довжини до 64 включно, що також дозволить оцінювати подібність доволі довгих двійкових послідовностей.

2 Гешування кожного з отриманих k-мерів

У якості геш-функції у нашому дослідженні буде використовуватися геш-функція Купина. Найменшим виходом Купини є 32 байтове геш-значення, тому для можливості представлення набору гешів у вигляді 64-бітних чисел взято тільки перші 8 байтів отриманих гешів. Такий підхід є можливим, оскільки криптографічні геш-функції розроблені таким чином, що є можливим зрізання вихідних даних до певного розміру, і зрізана геш-функція залишається безпечною криптографічною геш-функцією [13]. Зрізання також не є критичним з тієї причини, що для алгоритму MinHash не вимагається криптографічно стійка геш-функція.

3 Сортування отриманих гешів

4 Вибір невеликої кількості найменших геш-значень (скетчу), які і будуть представленням послідовності. Чим більш схожими є послідовності, тим більше MinHash вони ділитимуть між собою.

Для оцінки схожості цих двох наборів пропонується використовувати MinHash оцінку Жакарра. Для наборів k-мерів A та B алгоритм MinHash оцінює індекс Жакарра наступним чином:

$$jaccard(A_s, B_s) = \frac{|A_s \cap B_s|}{s}, \tag{3}$$

де A_s, B_s – підмножини такі, що $|A_s \cup B_s|$ дорівнює розміру скетчу, s. Розмір скетчу відповідає кількості MinHash, які зберігаються. Більші скетчі краще відображають послідовність, але за рахунок більшого розміру файлів та довшого часу порівняння.

Похибка оцінки MinHash відстані для заданого розміру скетчу s дорівнює $\sqrt{\frac{1}{s}}$ [11].

Після отримання індексу Жакарра MinHash відстань оцінюється за формулою

$$\frac{-\log(2.0 \times jaccard)/(1.0 + jaccard)}{k}, \tag{4}$$

де k – обраний розмір k-мерів. Такий метод дозволяє доволі швидко і з розумними затратами продуктивності оцінити схожість двох послідовностей ДНК з доволі точним наближенням.

Точність описаних методів порівняння послідовностей оцінюється у п. 4 даної роботи.

3. Аналіз статистичних властивостей випадкових та псевдовипадкових послідовностей на основі ДНК

3.1. Статистичне тестування псевдовипадкових послідовностей

У даному пункті надані результати дослідження послідовностей, отриманих шляхом виконання алгоритму 1 (п. 2.1). Дослідження проводилося з використанням набору тестів NIST STS [14]. Цей набір тестів для тестування генераторів випадкових чи псевдовипадкових чисел дає змогу з високою часткою ймовірності судити про те, наскільки послідовність, що генерується досліджуваним примітивом, є статистично безпечною, оскільки він включає у себе найбільш розповсюджені тести, які охоплюють більшість аспектів випадковості послідовності. Довжина послідовностей 13 МБ.

Таблиця 2

Результати статистичного тестування псевдовипадкових послідовностей на виході ДСТУ 7624:2014

Довжина ключа	Послідовність	Кількість тестів, у яких тестування пройшли більш 96 % послідовностей	Кількість тестів, у яких тестування пройшли більш 99 % послідовностей	Кількість тестів для яких значення P<0.001	Успіх
128	HS_Kalyna128Encrypted	188 (99 %)	143 (76 %)	1	+
	EC_Kalyna128Encrypted	188 (99 %)	127 (67 %)	0	+
	FC_Kalyna128Encrypted	188 (99 %)	134 (70 %)	0	+
	VO_Kalyna128Encrypted	189 (100 %)	138 (73 %)	1	+
	MV_Kalyna128Encrypted	188 (99 %)	131 (69 %)	0	+
	AP_Kalyna128Encrypted	187 (99 %)	130 (69 %)	0	+

Як видно з результатів дослідження, всі згенеровані послідовності успішно пройшли статистичне тестування. Найкращий результат для більш жорсткого критерію показала послідовність HS_Kalyna128Encrypted (з ДНК людини), для неї успішно пройдено 143 тести (76%). Для більш послабленого критерію найкращою є послідовність

VO_Kalyna128Encrypted (з ДНК калини), оскільки проходить всі 189 тестів (100 %). На рис. 4 представлено статистичні портрети досліджених послідовностей.

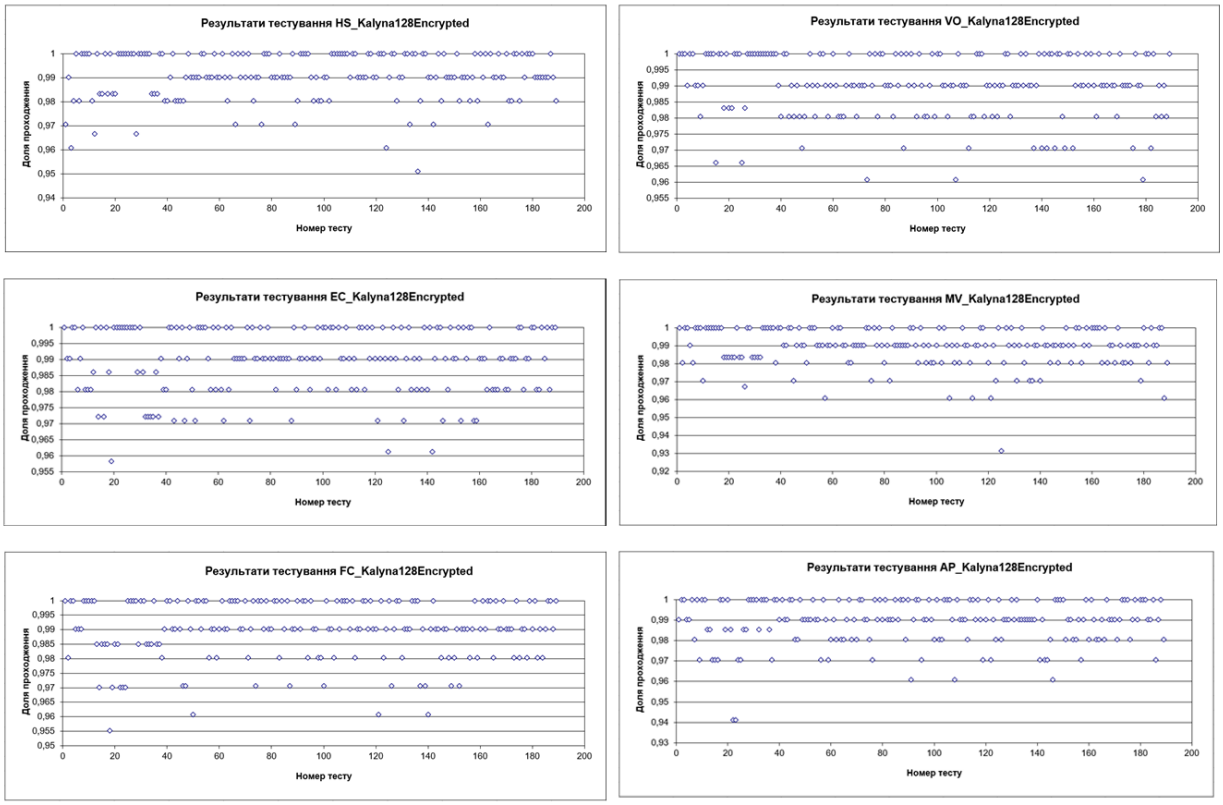


Рис. 4. Статистичні портрети псевдовипадкових послідовностей довжини 13 МБ

Результати тестування з використанням набору NIST STS показують, що псевдовипадкові послідовності з ДНК на виході алгоритму Калина мають гарні статистичні характеристики.

Оскільки результати для різних ДНК послідовностей є доволі схожими, у подальших дослідженнях будуть використовуватися тільки послідовності з ДНК людини.

Також для перевірки можливості генерування ПВП більшої довжини було отримано три послідовності довжини 151739136 байтів (144 МБ) та одну 1517391360 байтів (1,41 ГБ). Частину ключа зашифрування та синхропосилки для кожного з випадків показано у табл. 3.

Таблиця 3

Ключ зашифрування та синхропосилка для генерування ПВП

Генерування послідовностей довжини 13 МБ, першої послідовності довжини 144 МБ та довжини 1,41 ГБ	
Ключ	0x7E, 0x81, …, 0xD1, 0xB6
Синхропосилка	0xF3, 0x42, …, 0xD2, 0xFC
Генерування другої послідовності довжини 144 МБ	
Ключ	0x92, 0x48, …, 0xBC, 0x74
Синхропосилка	0x8A, 0x04, …, 0x2C, 0xDE
Генерування третьої послідовності довжини 144 МБ	
Ключ	0xFF, 0x06, …, 0x02, 0x4D
Синхропосилка	0x73, 0xA4, …, 0xFE, 0x3F

Дані значення є частиною послідовності, отриманої з квантового генератора.

Послідовності довжини 151739136 байтів вдалося успішно та доволі швидко протестувати з використанням NIST STS. Результати тестування у вигляді статистичних портретів надаються на рис. 5.

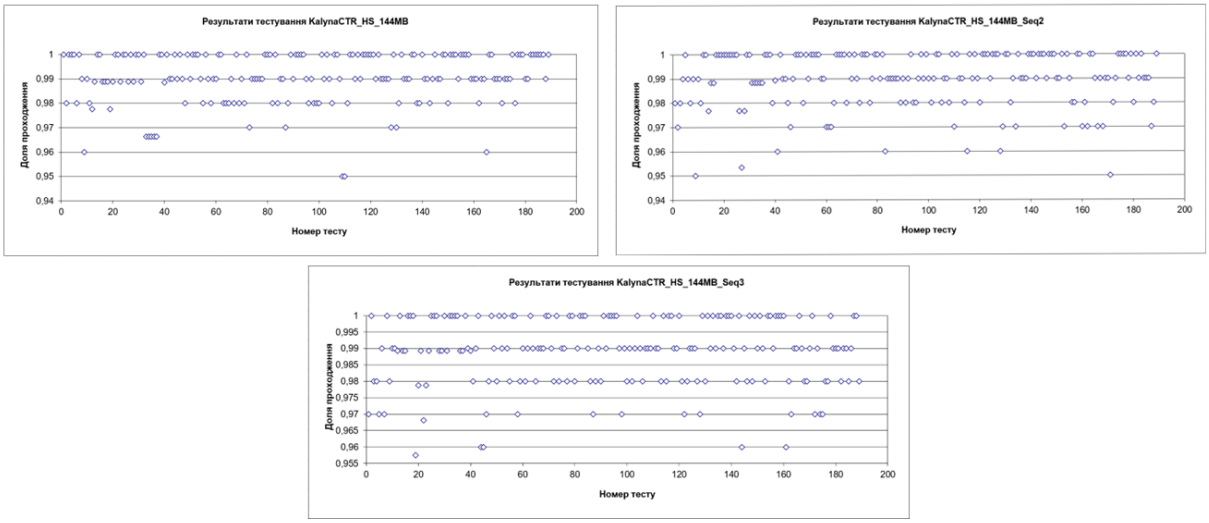


Рис. 5. Статистичні портрети псевдовипадкових послідовностей довжини 144 MB

Отримані з використанням NIST STS статистичні портрети також показують, що послідовності успішно проходять майже всі тести.

Додатково було виконано статистичне тестування послідовності довжини 1,41 ГБ. Тестування такої послідовності відбувалося майже 24 години, тому таке дослідження є доволі часомістким. Зауважте, що у статистичному портреті упущено значення для тесту «Перевірки шаблонів, що перекриваються», оскільки значення тільки для цього тесту є незадовільним (56/100) і псує наочність представлення результатів інших тестів (рис. 6).

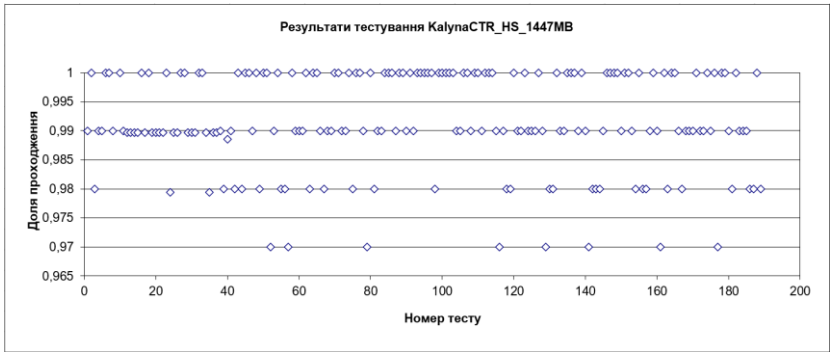


Рис. 6. Статистичний портрет послідовності KalynaCTR_HS_1447MB
(без урахування тесту шаблонів, що перекриваються)

Провал тесту «Перевірки шаблонів, що перекриваються» може свідчити про велику кількість *m*-бітних серій з одиниць у послідовності. Проте, це також може вказувати, що такі тести не розраховані на настільки довгі послідовності. Це перевірено у наступному пункті.

З отриманих результатів можна зробити висновок, що навіть досить довгі послідовності на виході ДСТУ 7624:2014 мають хороші статистичні характеристики. Це вказує на те, що такі послідовності можуть бути використані у якості ПВП для необхідних задач.

3.2. Статистичне тестування випадкових послідовностей

У даному пункті надано результати дослідження послідовностей, отриманих з використанням алгоритму 2 (п. 2.2).

За допомогою екстрактора було отримано три послідовності довжини 150994944 байтів (144 МБ) та одну 1509949440 байтів (1,40 ГБ). У якості основи було взято частину двійкової послідовності з ДНК людини. Ключ зашифрування та синхропосилка для кожної ітерації зашифрування отримувалася з використанням NPTRNG /dev/random [6] для ОС Linux. Тобто ключ та nonce при кожному повторному зашифруванні є випадковими даними, що і передбачено стандартом ДСТУ 7624:2014 для генерації випадкових послідовностей.

Результати аналізу властивостей у вигляді статистичних портретів надаються далі (рис. 7):

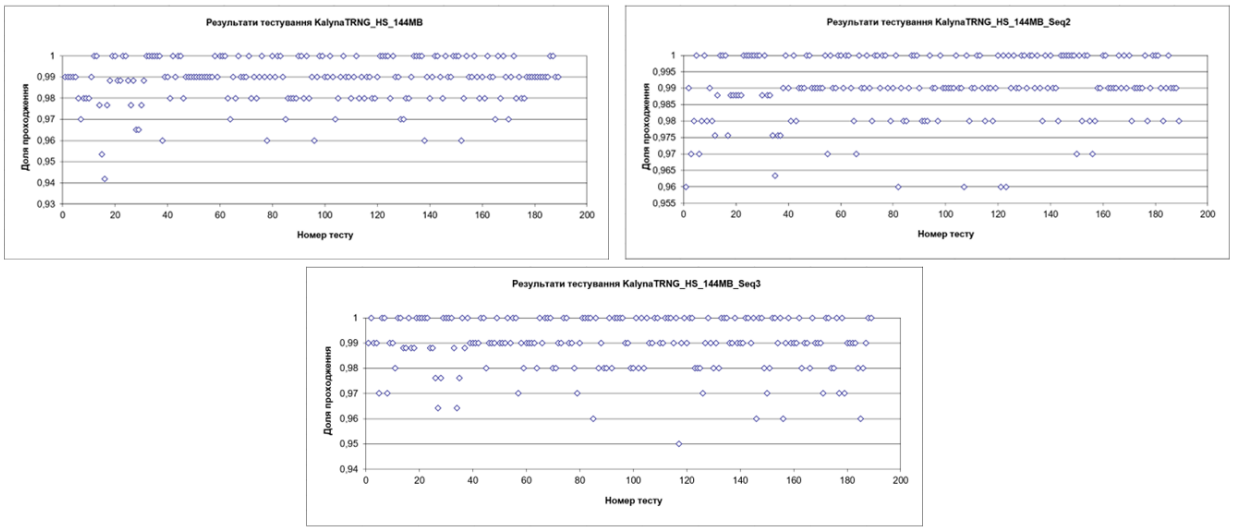


Рис. 7. Статистичні портрети випадкових послідовностей довжини 144 МВ

Отримані статистичні портрети підтверджують, що послідовності успішно проходять тестування.

Для послідовності довжини 1.4 ГБ тест «Перевірки шаблонів, що перекриваються» також показав незадовільний результат (55/100), проте всі інші тести було успішно пройдено. Для представлення у вигляді статистичного портрету значення проваленого тесту також буде упущено (рис. 8):

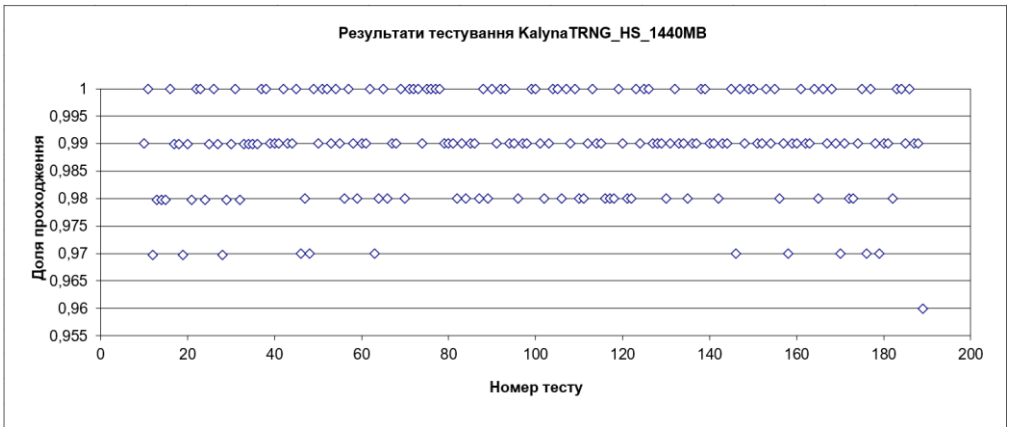


Рис. 8. Статистичний портрет послідовності KalynaTRNG_HS_1447 МВ
(без урахування тесту шаблонів, що перекриваються)

Для перевірки припущення про помилковість цього тесту для дуже довгих послідовностей було додатково перевірено послідовності великої довжини з обґрунтованих генераторів, а саме з /dev/random та CryptGenRandom [15]. Дослідження показало результати тесту «Перевірки шаблонів, що перекриваються» 59/100 та 54/100 відповідно. Хоча, як і для послідовності з екстрактора, всі інші тести було пройдено успішно. Це може означати, що саме даний тест не підходить для перевірки послідовностей таких довжин, а отже його результат не слід враховувати при оцінці.

Випадкові послідовності на виході екстрактора також майже повністю проходять всі тести з набору NIST STS. Це вказує на те, що послідовності з екстрактора на основі ДСТУ 7624:2014 можуть бути використані у якості ВП для необхідних задач.

У наступному пункті оцінено подібність послідовностей: як ДНК, так і двійкових.

4. Аналіз подібності ПВП, ВП та ДНК послідовностей для різних організмів

Для оцінки подібності пропонуються два розроблених методи на основі вимірювання k-мер відстані та MinHash відстані (пп. 2.3 та 2.4). Ці методи підходять як для оцінки ДНК послідовностей, які у даному дослідженні виступають у якості ДШ, так і двійкових послідовностей, на виході екстракторів.

4.1. Метод підрахунку k-мерів та k-мер відстані

4.1.1. Порівняння послідовностей ДНК

Для демонстрації методу виконано порівняння послідовностей ДНК для різних організмів з використанням алгоритму 3 (п. 2.3). Також перевірено двійкові послідовності, отримані на виході ДСТУ 7624:2014 на схожість з використанням цього алгоритму. Для перевірки обрано наступні параметри: k=5, як визначено за замовчуванням у пакеті R, також обрано k=7, 9 та 11, оскільки для підрахунку краще використовувати k-мери непарної довжини. Додатково було виконано перевірку на великій довжині k-мерів – k=31. Зверніть увагу, що схожість послідовностей визначається як (1–k-мер відстань) (табл. 4).

Таблиця 4

Результат обчислення k-мер відстані для послідовностей ДНК

Ген GNRHR2						
Послідовність 1	Послідовність 2	k-мер відстань				
		k=5	k=7	k=9	k=11	k=31
Послідовність з ДНК людини	Послідовність з ДНК шимпанзе	0.008139	0.020754	0.0288010	0.0362676	0.1056564
Послідовність з ДНК людини	Послідовність з ДНК макаки резус	0.019566	0.104332	0.1654499	0.2033711	0.4711847

Як видно з отриманих результатів, метод справді дозволяє з доволі точним наближенням оцінити схожість ДНК послідовностей. Найкращі результати метод показує саме у проміжку між розрахованими розмірами k-мерів 7 – 11. Отже, розрахунок розміру k-мерів відповідно до розміру послідовності і алфавіту є необхідним, чим більш точним є цей розрахунок, тим більш точною буде оцінка відстані.

Оскільки такий алгоритм можливо застосувати до послідовностей з двійковим алфавітом, далі перевірено послідовності різної довжини, отримані на виході ДСТУ 7624:2014 на подібність.

4.1.2. Порівняння ВП з екстрактора

Для оцінки подібності ВП на виході екстрактора згенеровано по дві послідовності різних довжин: 2, 5 та 13 МБ. Для генерування кожної з двох послідовностей у парі було використано різні частини послідовності ДНК людини. Довжини обрано таким чином, щоб довжиною k-мерів для їх оцінки було 31, 33 та 35 відповідно.

Кожну з пар послідовностей буде перевірено з використанням такої довжини, а також з довжинами, що є на крок меншими та більшими – 29 та 37.

У табл. 5 наведено результати оцінки k-мер відстані між згенерованими послідовностями.

Таблиця 5

Результат обчислення k-мер відстані для двійкових ВП з екстрактора

П1	П2	k-мер відстань				
		k=29	k=31	k=33	k=35	k=37
Kalyna_2MB_Seq1	Kalyna_2MB_Seq2	0.884817	**0.967398**	0.991574	0.997882	0.999470
Kalyna_5MB_Seq1	Kalyna_5MB_Seq1	0.763548	0.923910	**0.979371**	0.994756	0.998689
Kalyna_13MB_Seq1	Kalyna_13MB_Seq1	0.572017	0.828964	0.948761	**0.986419**	0.996545

Як видно з отриманих результатів, при використанні найбільш відповідної для довжини послідовностей довжини k-мерів (виділено жирним) – результати подібності двійкових

послідовностей є близькими до 1 – 3 %. Такий показник вказує на те, що послідовності на виході алгоритму Калина (ДСТУ 7624:2014) у режимі лічильника (гамування) є рівномірними та не схожими одна на одну (рис. 9).

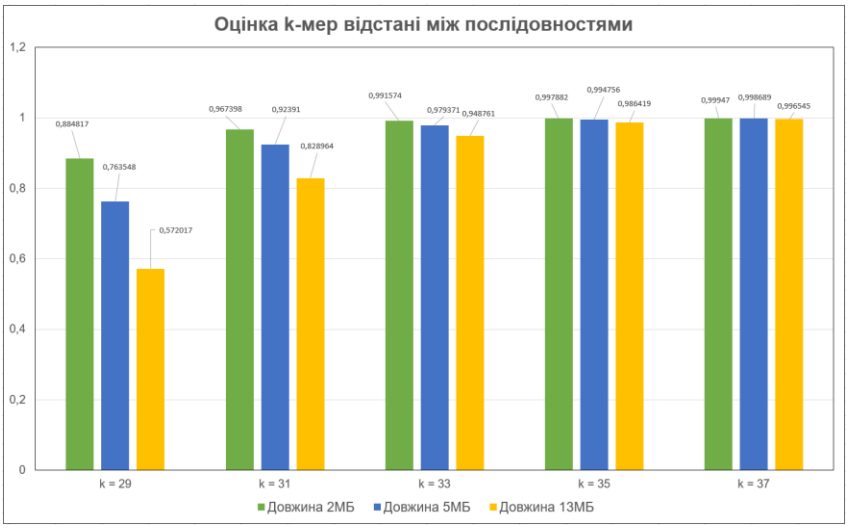


Рис. 9. Оцінка k-мер відстані між двійковими послідовностями на виході ДСТУ 7624:2014

У наступному пункті представлено покращений метод на основі MinHash з використанням у якості геш-функції алгоритму Купина (ДСТУ 7564:2014) [11].

4.2. Метод на основі MinHash відстані з використанням ДСТУ 7564:2014

4.2.1. Порівняння послідовностей ДНК

В табл. 6 наведено результати обчислення MinHash відстані для послідовностей ДНК різних організмів з алгоритмом 4 (п. 2.4).

Таблиця 6

Результат обчислення MinHash відстані для послідовностей ДНК

Ген GNRHR2					
Послідовність 1	Послідовність 2	MinHash відстань			
		k=7	*k*=9	*k*=11	*k*=31
Послідовність з ДНК людини	Послідовність з ДНК шимпанзе	0.0159405	0.0166228	0.0155702	0.0105054
Послідовність з ДНК людини	Послідовність з ДНК макаки резус	0.0583056	0.0650214	0.0612689	0.0473965

З отриманих результатів можна бачити, що оцінка, починаючи з k=11, є доволі точною і відповідає результатам, які можна отримати іншими інструментами порівняння ДНК, в тому числі і з вирівнюванням. Отримані результати вказують на те, що алгоритм дозволяє з достатньо хорошим наближенням оцінити подібність двох ДНК послідовностей. У наступному пункті виконано оцінку двійкових ВП з екстрактора таким самим методом.

4.2.2. Порівняння ВП з екстрактора

Для оцінки подібності ВП на виході екстрактора були взяті ті самі послідовності, що і у підп. 4.1.2. У табл. 7 наведено результати оцінки MinHash відстані між згенерованими послідовностями.

Як видно з отриманих результатів, при досягненні необхідного розміру k-мерів для відповідної довжини послідовності, відстань між послідовностями оцінюється у приблизно 15 – 16 %, при цьому показник подібності за індексом Жаккара є майже нульовим (0,001 – 0,005). Це означає, що такий метод для двійкового алфавіту може бути недостатньо чутливим. Проте при перевищенні розміру k-мерів на 1 – відстань оцінюється уже в 100 %, що означає повну відмінність між скетчами послідовностей (набором їхніх мінімальних гешів).

Таблиця 7

Результат обчислення MinHash відстані для двійкових послідовностей

П1	П2	MinHash відстань (попередньо порахований індекс Жакарра)				
		k=29	*k*=31	*k*=33	*k*=35	*k*=37
Kalyna_2MB_Seq1	Kalyna_2MB_Seq2	0.129022 (0.012)	**0.152819 (0.0044)**	0.195108 (0.0008)	1.00 (0.00)	1.00 (0.00)
Kalyna_5MB_Seq1	Kalyna_5MB_Seq1	0.0980334 (0.03)	0.123473 (0.011)	**0.155121 (0.003)**	0.177589 (0.001)	1.00 (0.00)
Kalyna_13MB_Seq1	Kalyna_13MB_Seq1	0.0674871 (0.076)	0.0974331 (0.025)	0.146434 (0.004)	**0.177589 (0.001)**	1.00 (0.00)

Отже, представлені методи порівняння є значно простішими у реалізації, ніж попередньо досліджені методи з використанням вирівнювання, і здатні забезпечувати оцінку схожості послідовностей, як ДНК, так і двійкових, з достатньо точним наближенням, майже ідентичним до методів на основі вирівнювання. Такі методи мають значну перевагу у продуктивності, забезпечують кращий час виконання. Також завдяки покращенню розрахунку k-мер відстані з використанням алгоритму MinHash та геш-функції Купина вдалося досягти додаткового прискорення швидкодії та значної економії апаратних ресурсів комп'ютера, зокрема, оперативної пам'яті. Проте такий метод може мати дещо низьку чутливість при дослідженні двійкових послідовностей, які завідома мають бути несхожими, наприклад, походячи з довірених PTRNG чи NPTRNG, а отже потребує подальшого дослідження та більш точного налаштування.

Висновки

1. Попередні дослідження вказують на теоретичну можливість використання у якості нефізичного джерела шуму та відповідно ГВП – ДНК. Як показав аналіз наукових джерел, детальних досліджень в цьому напрямі немає чи вони недоступні. Для оцінки та порівняння ПВП та ВП на основі використання ДНК потрібно провести широкі практичні та теоретичні дослідження з використанням ентропійних (стохастичних) методів та методик, а також визнаних статистичних методик (бажано стандартизованих).

2. ДНК можливо подати у вигляді послідовності, яка складається з алфавіту, у якому можливі чотири значення – A, C, G та T. Така послідовність може бути представлена у вигляді двійкової послідовності шляхом заміни цих значень на відповідні двійкові комбінації – A=00, C=01, G=10 та T=11. Наприклад, маючи таку послідовність як на рис. 1 – CAGGATCGTAGA – отримаємо двійкову послідовність 010010100011011011001000.

3. Для проведення досліджень було обрано декілька зразків ДНК різних організмів. Обрані послідовності ДНК було представлено у вигляді двійкових, як описано вище. Для кожної з них було проведено статистичне та стохастичне тестування. Отримані результати надають інформацію про те чи мають «сирі» послідовності достатньо випадковості і чи потребують вони покращення.

4. Статистичні характеристики «сирих» послідовностей не є задовільним, майже всі статистичні тести провалені. Хоча більшість тестів мінімальної ентропії мають незадовільне значення, а показники ентропії Шеннона та колізійної ентропії є хорошими, що вказує на присутність певної кількості ентропії у таких послідовностях, це дає можливість застосування екстракції випадковості з використанням, наприклад, БСШ.

5. Для отримання гарних ПВП та ВП на основі ДНК можливо використати екстрактор на основі ДСТУ 7624:2014 з довжиною ключа 256 біт у режимі гамування (CTR), оскільки саме такий режим пропонується використовувати з блоковим шифром для отримання ПВП та ВП у AIS 20/31.

6. Дослідження проводилося з використанням набору тестів NIST STS. Цей набір тестів для тестування генераторів випадкових чи псевдовипадкових послідовностей дає можливість з високою часткою ймовірності судити про те, наскільки послідовність, що генерується досліджуваним примітивом, є статистично безпечною, оскільки він включає у себе найбільш

розповсюджені тести, які охоплюють більшість аспектів випадковості послідовності. Довжина таких послідовностей може бути 13 МБ.

7. Як видно з рис. 4, всі згенеровані послідовності успішно пройшли статистичне тестування. Найкращий результат для більш жорсткого критерію показала послідовність HS_Kalyna128Encrypted (з ДНК людини), для неї успішно пройдено 143 тести (76 %). Для більш послабленого критерію найкращою є послідовність VO_Kalyna128Encrypted (з ДНК калини), оскільки проходить всі 189 тестів (100 %).

8. За допомогою екстрактора було отримано три послідовності довжини 150994944 байтів (144 МБ) та одну 1509949440 байтів (1,40 ГБ). У якості основи було взято частину двійкової послідовності з ДНК людини. Ключ зашифрування та синхропосилка для кожної ітерації зашифрування отримувалася з використанням NPTRNG /dev/random [6] для ОС Linux.

9. Для оцінки подібності ВП на виході екстрактора згенеровано по 2 послідовності різних довжин: 2, 5 та 13 МБ. Для генерування кожної з двох послідовностей у парі було використано різні частини послідовності ДНК людини. Довжини обрано таким чином, щоб довжиною k-мерів для їх оцінки було 31, 33 та 35 відповідно.

10. Для порівняння двох ДНК послідовностей для кожної з них спочатку підраховується кількість k-мерів (підрядків) у її складі. Створюється таблиця, де кожному з k-мерів (4^k для стандартного алфавіту ДНК і, наприклад, 2^k у випадку двійкового алфавіту) відповідає кількість входжень конкретного k-меру у послідовність, що досліджується.

11. При досягненні необхідного розміру k-мерів для відповідної довжини послідовності, відстань між послідовностями оцінюється у приблизно 15 – 16 %, при цьому показник подібності за індексом Жаккара є майже нульовим (0,001 – 0,005). Це означає, що такий метод для двійкового алфавіту може бути недостатньо чутливим. Проте, при перевищенні розміру k-мерів на 1 сходинку вище – відстань оцінюється уже в 100.

12. Представлені методи порівняння ПВП та ВП ДНК значно простіші у реалізації, ніж попередньо досліджені методи з використанням вирівнювання, і здатні забезпечувати оцінку схожості послідовностей як ДНК, так і двійкових, з достатньо точним наближенням, майже ідентичним до методів на основі вирівнювання. Такі методи також мають значну перевагу у продуктивності, забезпечують кращий час виконання.

13. Таким чином, оцінка подібності послідовностей з використанням k-мер відстані показує, що при відповідних до довжини послідовності значеннях *k* подібність між послідовностями на виході екстрактора є мінімальною.

14. У цілому вважаємо, що актуальними та необхідними є подальші дослідження, оцінка та порівняння різних ДНК, встановлення подібності ПВП різних ДНК тощо.

Список літератури:

1. Matthias Peter, Werner Schindler. A Proposal for Functionality Classes for Random Number Generators. Version 2.36 – Current intermediate document for the AIS 20/31 workshop. 2023. URL: https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Certification/Interpretations/AIS_31_Functionality_classes_for_random_number_generators_e_2023.html?nn=910324
2. Горбенко І. Д., Дерев'янко Я. А., Горбенко Д. Ю. ДНК – джерело шуму та нефізичних випадкових послідовностей. Основні положення практичних досліджень. 2024. URL: https://cyberwarfare.viti.edu.ua/assets/files/Cyberwarfare_2024.pdf
3. Національна Бібліотека Медицини США. URL: https://ftp.ncbi.nlm.nih.gov/genbank/
4. ДНК банк Японії. URL: https://ddbj.nig.ac.jp/arsa/quick_search?lang=en;jsessionid=5FC7A34BD7FA0826284A6619E85BCA97
5. ДСТУ 7624:2014. Інформаційні технології. Криптографічний захист інформації. Алгоритм симетричного блокового перетворення. 2015. URLhttps://online.budstandart.com/ua/catalog/doc-page.html?id_doc=109736
6. Linux manual page. random. URL: https://man7.org/linux/man-pages/man4/random.4.html
7. Shaun Wilkinson. Introduction to the kmer R package. 2019. URL: https://cran.r-project.org/web/packages/kmer/vignettes/kmer-vignette.html.
8. Edgar, R. C. Local homology recognition and distance measures in linear time using compressed amino acid alphabets. 2004. URL: https://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC373290&blobtype=pdf
9. Mash. Fast genome and metagenome distance estimation using MinHash. Sketches. URL: https://mash.readthedocs.io/en/latest/sketches.html

10. Andrei Z. Broder. On the resemblance and containment of documents. 1998. URL: https://web.archive.org/web/20150131043133/http://gatekeeper.dec.com/ftp/pub/dec/SRC/publications/broder/positano-final-wpnums.pdf

11. Brian D. Ondov, Todd J. Treangen, Páll Melsted, Adam B. Mallonee, Nicholas H. Bergman, Sergey Koren & Adam M. Phillippy. Mash: fast genome and metagenome distance estimation using MinHash. 2016. URL: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-0997-x

12. ДСТУ 7564:2014. Інформаційні технології. Криптографічний захист інформації. Функція гешування. 2015. URL: https://usts.kiev.ua/wp-content/uploads/2020/07/dstu-7564-2014.pdf.

13. John Kelsey. Truncation Mode for SHA. 2005. URL: https://csrc.nist.gov/csrc/media/events/first-cryptographic-hash-workshop/documents/kelsey_truncation.pdf

14. NIST SP 800-22 Rev. 1. A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications. 2010, URL: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-22r1a.pdf

15. Microsoft Documentation. CryptGenRandom function (wincrypt.h). 2021. URL: https://learn.microsoft.com/en-us/windows/win32/api/wincrypt/nf-wincrypt-cryptgenrandom

Надійшла до редколегії 10.06.2024

Відомості про авторів:

Дерев'янко Ярослав Андрійович – АТ «Інститут інформаційних технологій», науковий співробітник-консультант; Україна; e-mail: yarik0009258@gmail.com; ORCID: https://orcid.org/0000-0002-3290-3373

Єсіна Марина Віталіївна – канд. техн. наук, доцент, Харківський національний університет імені В. Н. Каразіна, доцент кафедри безпеки інформаційних систем і технологій, факультет комп'ютерних наук, АТ «Інститут Інформаційних Технологій», науковий співробітник-консультант; Україна; e-mail: m.v.yesina@karazin.ua; ORCID: https://orcid.org/0000-0002-1252-7606

Горбенко Дмитро Юрійович – Харківського національного університету імені В. Н. Каразіна, студент факультету комп'ютерних наук, АТ "Інститут Інформаційних Технологій", молодший інженер-програміст; Україна; e-mail: jsciitua@gmail.com

УДК 004.05 **DOI:10.30837/rt.2024.2.217.03**

В.І. ЄСІН, д-р техн. наук, В.В. ВІЛІГУРА, Д.Ю. УЗЛОВ, канд. техн. наук

ОГЛЯД ІСНУЮЧИХ МОДЕЛЕЙ ТА ОСНОВНИХ ПРИНЦИПІВ НУЛЬОВОЇ ДОВІРИ

Вступ

Виклики безпеки XXI століття характеризуються змінами та непередбачуваністю [1]. Диджиталізація (digitalization) нашого світу веде до зростання кількості комунікацій. Останні тенденції, такі як хмарні обчислення, Інтернет речей та використання концепцій CYOD та BYOD призводять до збільшення розмірів існуючих мереж. Все більше і більше пристроїв та сервісів обмінюються інформацією всередині корпоративних мереж, а також за їх межами. Ці зміни призводять до появи нових складних вимог до мережевої безпеки, яким існуючі рішення слабо відповідають [2], про що свідчить зростання кількості витоків даних та хакерських атак (зломи стають все масштабнішими та сміливішими, зачіпаючи все: від баз даних клієнтів та громадян до даних про вакцини та маршрутизатори Wi-Fi [3]). Як відомо, традиційна мережева безпека фокусується на захисті периметра. Тобто більшість концепцій мережевої безпеки ґрунтуються на розподілі внутрішніх та зовнішніх мереж. Всі користувачі, пристрої та служби у захищеній внутрішній мережі вважаються довіреними, тоді як зовнішні користувачі, пристрої та служби класифікуються як ненадійні. Однак недоліки цього підходу стають очевидними, якщо врахувати, що зловмисники у разі компрометації суб'єктів (кінцевих користувачів, застосунків та інших нефізичних сутностей), які запитують інформацію та часто отримують широкий доступ до безлічі корпоративних ресурсів, шляхом видачі себе за іншу особу та ескалації (підвищення привілеїв) можуть отримати доступ до ресурсів усередині або за межами мережі. Більше того, багато підприємств (організацій, компаній) не мають чітко визначеного периметра. Периметр втрачає свою актуальність через кілька факторів, включаючи зростання хмарних обчислень, мобільність та зміни у сучасному штаті співробітників (використання віддалених працівників) [4]. Крім того, загрозу становлять і внутрішні зловмисники (інсайдери). Таким чином, ідея про те, що жодна мережа (ні внутрішня, ні зовнішня) не заслуговує на довіру, набирає обертів як у наукових колах, так і на практиці [2].

Щоб захистити сучасне цифрове підприємство, необхідна комплексна стратегія для безпечного доступу у будь-який час і в будь-якому місці до своїх корпоративних ресурсів (застосунків, застарілих/успадкованих систем, даних, пристроїв тощо) незалежно від того, де вони розташовані [5]. Дійсно, зростання хмарних обчислень, Інтернету речей, бізнес-партнерів та зростаючої кількості віддалених працівників підвищує складність захисту цифрових ресурсів підприємства, оскільки існує більше точок входу, виходу та доступу до даних, ніж будь-коли раніше, а існуючі рішення відчувають труднощі з реагуванням на динамічні зміни, оскільки вони часто складаються зі статичних наборів правил, міжмережевих екранів, VPN та підмереж. Внутрішня мережа захищена настільки, наскільки захищено найгірше захищений пристрій або застосунок. Заходів, що обмежують бічне переміщення (lateral movement) по мережі, або дуже мало, або вони взагалі відсутні. IP-адреси пристроїв та сервісів відомі ззовні. Існуючі рішення спочатку встановлюють з'єднання, а потім перевіряють права доступу. Це робить їх потенційними цілями, які можна використовувати для проникнення в мережу або порушення її роботи, наприклад, за допомогою розподілених атак типу «відмова в обслуговуванні» (DDoS). Файли журналів зберігаються на централізованих серверах журналів. Отримавши доступ до таких файлів, зловмисники можуть замаскувати свою діяльність та стерти сліди [2]. Тому підприємства сьогодні переосмислюють традиційний периметр безпеки мережі, схиляючись до нової концепції та архітектури захисту.

Такою концепцією стала сьогодні парадигма безпеки, що отримала назву «нульова довіра» (Zero Trust – ZT). За своєю суттю «нульова довіра» – це філософія, підхід та набір керівних принципів кібербезпеки, що використовуються для створення стратегії, яка фокусується

на переміщенні захисту мережі від широких статичних периметрів мережі до вужчого зосередження уваги на суб'єктах, активах підприємства (а саме, пристроях, компонентах інфраструктури, застосунках, віртуальних та хмарних компонентах) та окремих або невеликих групах ресурсів [4, 6, 7]. Нульова довіра – це не єдина архітектура, а набір керівних принципів для робочого процесу, проєктування системи та операцій, які можна використовувати для покращення стану безпеки будь-якої класифікації або рівня чутливості [7]. Основна ідея концепції нульової довіри полягає в тому, що не існує областей, які заслуговують на довіру. І її основний принцип «ніколи не довіряти, завжди перевіряти» [8]. Цей більш обмежувальний підхід спрямовано на поліпшення захисту ресурсів. Хоча «нульова довіра» починалася як вузькоспрямований підхід, який полягає в тому, щоб не довіряти жодним мережевим ідентифікаційним/обліковим даним (ідентифікаторам) доти, доки вони не будуть автентифіковані та авторизовані, його масштаби по праву розширилися, щоб забезпечити набагато ширший набір можливостей безпеки у середовищі підприємства (організації, компанії, корпорації). Архітектура нульової довіри (zero trust architecture – ZTA) враховує нові тенденції, приділяючи особливу увагу захисту ресурсів, а не периметру мережі, оскільки розташування мережі більше не розглядається як основний компонент забезпечення безпеки, необхідної для ресурсу [4, 7]. ZTA ніколи не надає доступ до ресурсів до тих пір, поки суб'єкт, актив (це елемент/об'єкт, що представляє цінність для зацікавлених сторін); актив може бути матеріальним (наприклад, фізичний об'єкт, такий як апаратне забезпечення, вбудоване програмне забезпечення, обчислювальна платформа, мережевий пристрій або інший технологічний компонент) або нематеріальним (наприклад, люди, дані, інформація, програмне забезпечення, можливості, функції, послуги, товарний знак, авторське право, патент, інтелектуальна власність, імідж чи репутація) [9]) або робоче навантаження не будуть верифіковані за допомогою процедур автентифікації та прав/дозволів на виконання певних дій (авторизації) [4].

Згідно з результатами дослідження, проведеного компанією Grand View Research, обсяг світового ринку систем безпеки з нульовою довірою в 2022 р. склав 24,84 мільярда доларів США, а в період з 2023 по 2030 р. очікується сукупний середньорічний темп зростання (CAGR – compound annual growth rate) на рівні 16,6 % (що за прогнозами дозволить вийти на цифру 82,45 мільярда доларів США до 2030 р.) [10, 11]. Gartner, Inc. прогнозує, що до 2026 р. 10 % великих підприємств матимуть зрілу та вимірну програму нульової довіри порівняно з менш ніж 1 % у 2023 р. [12].

Однак, незважаючи на популяризацію концепції нульової довіри та очевидні переваги у сфері безпеки від її застосування, на підприємствах виникають певні складнощі щодо її реалізації [13 – 15]. Якщо у 2021 р. згідно зі звітом компанії Fortinet [13], 40 % респондентів заявили, що їхня стратегія нульової довіри повністю реалізована, то у 2023 р. лише 28 % компаній оголосили про те, що вони вже мають повне рішення з нульовою довірою. Ці цифри показують, що, швидше за все, робота з впровадження концепції нульової довіри виявилася трохи складнішою, ніж передбачалося. Деякі проблеми стали очевидними лише після того, як кілька рішень вже були впроваджені, і виникла потреба у взаємодії між розрізненими рішеннями. Основними серйозними факторами, що перешкоджають, за даними того ж звіту, стали брак інформації для вибору рішення з нульовою довірою (на нього вказали 16 % організацій, у тому числі 24 % серед невеликих компаній) та відсутність кваліфікованих розробників/постачальників (на нього вказали 24 % організацій). 4 % організацій з опитаних вказали на нестачу людських ресурсів. Ще одним важливим висновком цього звіту є те, що розгортання рішень від кількох постачальників створює нові проблеми для організацій, включаючи ненавмисне створення проблем з безпекою та високі операційні витрати через розростання постачальників і рішень (на відсутність необхідних бюджетних коштів для проведення IT-змін прямо зараз – вказали 17 % організацій).

Як видно з викладеного, існує проблема, пов'язана з певним дефіцитом поінформованості про підхід нульової довіри (про його теоретичний та практичний потенціал) для вибору правильного рішення. Стаття націлена на вирішення цієї проблеми шляхом узагальнення

наявних досліджень та досвіду різних міжнародних компаній, які впроваджують даний підхід на практиці. У стислому викладі розглядаються моделі та ключові принципи нульової довіри, запропоновані відомими міжнародними організаціями та компаніями, які допоможуть розібратися у фундаментальному зрушенні у підході до інформаційної безпеки, кібербезпеки. Дослідження, що проводились у цій роботі, спираються на публікації різних міжнародних авторитетних видань (у тому числі стандартів та деяких урядових організацій по всьому світу), присвячених концепції нульової довіри.

1. Історія та еволюція концепції нульової довіри

Концепція нульової довіри стала відомою в кібербезпеці ще до появи терміну «нульова довіра», який був представлений у звіті [16] відомої американської дослідницької та консалтингової компанії Forrester (одного з ключових дослідників ринків інформаційних технологій). Ідея концепції, згідно з якою жодному учаснику мережі (як внутрішньої, так і зовнішньої) не можна довіряти, а будь-який доступ до ресурсів підприємства є потенційною загрозою, виникла ще на початку розвитку безпечних обчислень [2]. Так, ще 1975 р. автори роботи [17] пропонували повне посередництво у доступі та мінімальні привілеї. У 2007 р. Агентство оборонних інформаційних систем (DISA – Defense Information Systems Agency) та Міністерство оборони (Department of Defense – DoD) США опублікували свою роботу щодо більш безпечної корпоративної стратегії під назвою «чорне ядро» («black core») [1]. Чорне ядро передбачало перехід від моделі безпеки на основі периметра до такої, яка зосереджена на безпеці окремих транзакцій. В результаті роботи Jericho Forum міжнародною групою в галузі безпеки було оприлюднено ідею деперіметризації (de-perimeterization) [18] – обмеження неявної довіри на основі розташування мережі та обмеження покладатися на єдиний статичний захист у великому сегменті мережі [7]. Далі концепції деперіметризації розвивалися і вдосконалювалися, перетворюючись на ширшу концепцію, а саме концепцію нульової довіри. У звіті Forrester [16] відображено ідеї, які обговорювалися в галузі протягом кількох років, зокрема, при нульовій довірі весь мережевий трафік не є довіреним. Тобто фахівці з безпеки повинні: перевіряти та захищати всі ресурси; обмежувати та суворо забезпечувати контроль доступу; перевіряти та реєструвати весь мережевий трафік. Згодом у компанії Forrester розвинули концепцію нульової довіри, у ту, що зараз відома, як *розширення «нульової довіри»* (Zero Trust eXtended – ZTX) [19].

Приблизно в той же час Google розпочала свою реалізацію концепції комп'ютерної безпеки з нульовою довірою (BeyondCorp), яка переносить контроль доступу з традиційного периметру мережі на окремі пристрої та користувачів. Серія статей, починаючи з 2014 р., компанії Google [20 – 26] сприяли внесенню значного внеску до концепції комп'ютерної безпеки з нульовою довірою. BeyondCorp використовує низку політик безпеки, включаючи автентифікацію, авторизацію та контроль доступу, щоб гарантувати, що лише авторизовані користувачі можуть отримати доступ до корпоративних ресурсів. BeyondCorp складається з безлічі взаємодіючих компонентів, які гарантують, що лише пристрої та користувачі, що пройшли відповідну автентифікацію, мають право доступу до необхідних корпоративних застосунків [20].

У 2017 р. провідна світова дослідницька та консалтингова компанія у сфері інформаційних технологій Gartner, Inc. переглянула та оновила свою концепцію *безперервної адаптивної оцінки ризиків та довіри* (*Continuous Adaptive Risk and Trust Assessment* – CARTA), яка має багато загальних принципів із нульовою довірою. CARTA [27] надає не тільки елементи ідентифікації та даних, але також включає ризики та положення, пов'язані з ідентифікацією та пристроями, що мають доступ до середовища.

Завдяки публікації Національного інституту стандартів та технологій США (NIST) про архітектуру нульової довіри [7], а також пов'язаної з нею проєктом Національного центру передового досвіду в галузі кібербезпеки США (US National Cybersecurity Center of Excellence – NCCoE) [4], концепції нульової довіри у масштабах корпоративної мережевої безпеки

та безпеки даних підприємств стали приділяти більшої уваги. Федеральним агентствам вже понад десять років рекомендується перейти до безпеки, заснованої на принципах нульової довіри, створюючи можливості та політики, такі як Федеральний закон про модернізацію інформаційної безпеки (*Federal Information Security Modernization Act* – FISMA), якому підпорядковуються Система управління ризиками (*Risk Management Framework* – RMF); Федеральне управління ідентифікацією, обліковими даними та доступом (*Federal Identity, Credential, and Access Management* – FICAM); Довірені підключення до Інтернету (*Trusted Internet Connections* – TIC); і програми/системи безперервної діагностики та пом'якшення наслідків (*Continuous Diagnostics and Mitigation* – CDM). Усі ці програми спрямовані на обмеження доступу до даних і ресурсів для авторизованих сторін. Коли ці програми починалися, вони були обмежені технічними можливостями інформаційних систем. Політики безпеки були здебільшого статичними та застосовувалися у вузлових точках, які підприємство могло контролювати, щоб отримати найбільший ефект від вкладених зусиль. У міру розвитку технологій з'явилася можливість для безперервного детального аналізу та оцінки запитів на надання доступу (відповідно до принципу необхідності доступу), щоб знизити ризик втрати даних через злому облікових записів, атак зловмисників, що стежать за мережею та інших загроз [7]. І тепер уже відповідно до концепції нульової довіри забезпечується детальний контроль доступу з урахуванням ідентифікаційних та контекстно-залежних даних, що реалізується автоматично спеціалізованою системою.

Окремі принципи нульової довіри самі по собі не нові, новим є той факт, що всі ці принципи використовуються комплексно для захисту ресурсів підприємства [2]. При традиційному управлінні доступом права (привілеї) зазвичай призначаються заздалегідь з урахуванням посади (виконуваних функціональних обов'язків), а в рішеннях з нульовою довірою з'являється додатковий механізм – центр (пункт, точка) прийняття рішення про політику, що забезпечує динамічне прийняття рішень про доступ на основі наявних політик, а також вхідних даних із деяких важливих зовнішніх джерел. Якщо традиційні системи управління доступом розроблялися в основному з орієнтацією на користувача, то в епоху цифрової трансформації контроль доступу повинен також поширюватися на автономні системи та інтелектуальні агенти, що й робиться в рішеннях, створюваних на основі концепції нульової довіри. Слід зауважити, що, хоча автоматичні дії є основними в операційній діяльності в нових умовах, це не скасовує можливості використання у нових системах ручного втручання або включення в робочий процес конкретних дій вручну перед активацією автоматичного реагування. Крім того, сьогодні є деякі недостатньо ефективні (або вже застарілі) рішення (приклади деяких з них наведені в табл. 1), які вимагають заміни або модифікації, вдосконалення, адаптації до вимог систем, відповідним стратегії нульової довіри.

Таблиця 1

Існуючі рішення безпеки та їх недоліки

Існуючі рішення	**Недоліки рішення**
Перепустки	Втрата носія або його викрадення.
Сегментація за допомогою DNS-сервера	Занадто широкий доступ до мережі.
NAC (*Network Access Control*)	Коштовність та недостатня швидкість, відсутність можливості використання для хмарних середовищ.
IDS/IPS (*Intrusion Detection System / Intrusion Prevention System*)	Відсутність можливості застосування у хмарному середовищі, наявні помилкові спрацьовування при перевірці доступу.
VPN	Занадто проста процедура отримання доступу, надає широкий доступ до мережі.
SIEM (*Security Information* and *Event Management*)	Відсутність можливості керування віддаленим доступом.

Якщо порівнювати традиційну модель безпеки (безпеки периметра) і модель нульової довіри, то слід звернути на їхню принципову відмінність – модель нульової довіри не має

«зони довіри» і заснована на перевірці без довіри, навіть якщо це внутрішній користувач. У той час як модель безпеки периметра орієнтована на блокування, модель нульової довіри передбачає ретельну та постійну перевірку кожного користувача та пристрою, що намагається підключитися до ресурсів, незалежно від їхнього розташування [28]. Ці та інші аспекти характерних відмінностей даних моделей наведені в табл. 2.

Таблиця 2

Порівняння традиційної моделі безпеки та моделі нульової довіри

Характеристика	**Традиційна модель безпеки**	**Модель нульової довіри**
Підхід	Довіряй але перевіряй.	Нікому не довіряй і все перевіряй.
Межа довіри	Зовнішня (немає довіри – non trust). Внутрішня (довірена – trust).	Мікросегментація (мережі поділяються на дрібніші сегменти або безпечні зони, щоб обмежити бічне переміщення (lateral movement) загроз; кожен сегмент має свої засоби керування доступом користувачів та політики безпеки).
Мережева архітектура	Модель «замок та рів» (castle and moat) з підвищеним акцентом на захист периметра.	Децентралізована та мікросегментована, з детальним контролем доступу.
Контроль доступу	Контроль доступу на основі IP.	Керування доступом, орієнтоване на дані.
Автентифікація	Після перевірки при початковому доступі.	Перед доступом та постійна перевірка.
Керування безпекою	Індивідуальний моніторинг та видимість.	Видимість, автоматизація та оркестрування поведінки, пристроїв, сервісів/послуг та безпеки.
Політика безпеки	Заздалегідь встановлені правила та загальна політика.	Деталізовані правила та адаптивні політики (оцінка рівня безпеки).
Шифрування зв'язку	Зовнішня мережа (шифрування). Внутрішня (без шифрування).	Повне шифрування трафіку.
Реагування на порушення	Як тільки периметр порушено, зловмисники можуть здобути свободу дій.	Навіть якщо порушення сталося, переміщення зловмисників ретельно відстежуються.

В цілому ж, архітектура нульової довіри – це комплексний/наскрізний (end-to-end) підхід до корпоративних ресурсів і безпеки даних, який охоплює ідентифікацію фізичних та нефізичних осіб/сутностей, облікові дані, керування доступом, операції, кінцеві точки, середовища розміщення та інфраструктуру взаємозв’язку [7]. Інтеграція раніше розрізнених засобів забезпечення безпеки, систем інфраструктури та корпоративних систем має важливе значення. Інтеграція засобів ідентифікації та безпеки дозволяє створити комплексний механізм безпеки, за допомогою якого рішення нульової довіри можуть забезпечити безпечніше середовище.

При цьому слід зазначити, що концепція нульової довіри продовжує розвиватися в міру того, як постачальники та організації зі стандартизації переглядають та вдосконалюють специфікації та реалізації нульової довіри, визнаючи це фундаментальною зміною у підході до інформаційної безпеки [6].

2. Моделі нульової довіри

2.1. Розширена модель нульової довіри компанії Forrester

Компанія Forrester, як зазначалося вище, випустила свою першу модель нульової довіри у 2010 р., яку у наступні роки переглянула та внесла зміни. В результаті було створено так звану розширену модель нульової довіри (ZTX – Zero Trust eXtended) [19], в якій було виділено сім компонентів нульової довіри: п'ять для контролю безпеки та два – для взаємодії між компонентами (рис. 1). Розглянемо їх докладніше.

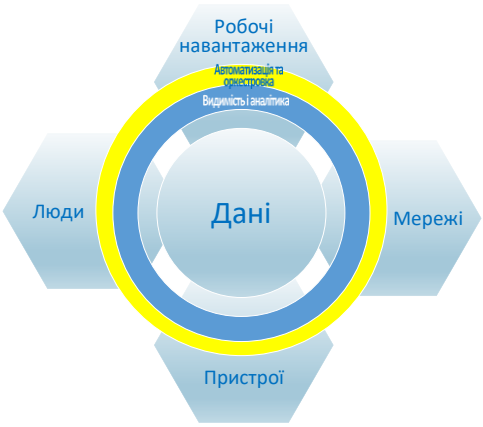


Рис. 1. Розширена модель нульової довіри компанії Forrester

Розширена модель нульової довіри надає багатший контент і всебічну модель, в якій дані розташовуються в центрі (рис. 1). При цьому фахівці Forrester загострюють увагу на тому, що різке збільшення обсягу даних як у локальних, так і в хмарних середовищах загострює проблему їх захисту, яку необхідно вирішувати та вирішувати її треба, виходячи з нових вимог та можливостей нових технологій та підходів.

Отже, *дані* (*data*) є центром моделі ZTX (рис. 1), а безпека даних є одним із стовпів стратегії нульової довіри. Захист даних та керування ними, категоризація та розробка схем класифікації даних, а також шифрування даних як при зберіганні, так і при передачі є ключовими елементами будь-якого підходу з нульовою довірою. Крім того, повинна бути система запобігання втраті даних (*Data Loss Prevention* – DLP), яка повинна бути частиною архітектури нульової довіри, а також бути пов'язаною з моделлю політики з можливістю застосування політик контекстного доступу, де тільки це можливо.

Оточуючі елементи – *робочі навантаження* (*workloads*), *мережі* (*networks*), *пристрої* (*devices*) і *люди* (*people*) – є провідниками даних і, отже, також потребують захисту.

Мережевий компонент (стовп) моделі ZTX в першу чергу орієнтований на сегментацію мережі (як на рівні користувачів, так і на рівні сервера) для забезпечення кращої безпеки на основі атрибутів, пов'язаних з ідентифікаційними даними. Здатність сегментувати, ізолювати та контролювати мережу є ключовим елементом управління для нульової довіри. При цьому слід розуміти, що підприємства сьогодні мають багато різних компонентів, що становлять традиційну інфраструктуру мережевої безпеки. Це і міжмережеві екрани наступного покоління (Next-Generation Firewalls – NGFW), і міжмережові екрани веб-застосунків (Web Application Firewalls – WAF), і рішення для контролю доступу до мережі (NAC), і системи захисту від вторгнень (Intrusion Protection Systems – IPS). Як правило, всі ці компоненти відіграють важливу роль і в рамках концепції нульової довіри.

Компонент (стовп) моделі ZTX «*Люди*». Останнім рубежем будь-якої стратегії нульової довіри є встановлення та суворе дотримання обмежень на доступ користувачів, а також забезпечення їхньої безпеки при взаємодії з Інтернет-ресурсами. З позиції Forrester, даний компонент повинен включати кілька елементів системи керування ідентифікацією і доступом (Identity and Access Management – IAM), в рамках якої відмінно себе зарекомендували такі добре вивчені моделі, як управління доступом на основі ролей (Role-Based Access Control – RBAC) та атрибутів (Attribute-Based Access Control – ABAC). При цьому технологія нульової довіри дозволяє використовувати дані моделі більш широко та ефективно у всій корпоративній інфраструктурі. Важливе значення для реалізації технології нульової довіри, як ключових елементів у рамках компонента «*Люди*», також мають багатофакторна автентифікація (Multi-Factor Authentication – MFA) та система єдиного входу (Single Sign On – SSO) з використанням сучасних відкритих стандартів, таких як OAuth та SAML.

Робочі навантаження – це термін високого рівня, який за визначенням Forrester, відноситься до всього стеку застосунків, від рівня застосунку до гіпервізора або автономних

компонентів обробки, таких як контейнери і віртуальні машини [19]. Ці застосунки слід розглядати як вектор загрози (особливе занепокоєння викликають робочі навантаження, що виконуються в публічних хмарах), і до них мають бути застосовані елементи керування та технології нульової довіри. Для нульової довіри потрібне управління доступом до робочого навантаження на основі метаданих, що послідовно застосовується в гібридних середовищах.

Пристрої з нульовою довірою. Технології Інтернету речей та мережевих пристроїв створили величезну сферу потенційного ризику для мереж та підприємств. Розумні (smart), мобільні пристрої сьогодні набули широкого поширення по всьому ринку товарів. Однак у результаті такого прориву виявився зворотний бік цього процесу – відкрилися нові можливості шляхів поширення коду та засобів, які фахівці безпеки повинні відстежувати та розглядати як ненадійні у будь-якій інфраструктурі. Щоб дійсно перейти до стратегії нульової довіри, фахівці з безпеки повинні мати можливість постійно ізолювати, захищати і контролювати кожен пристрій у мережі [19]. Іншими словами, модель безпеки для *Пристроїв* повинна включати ідентифікацію, облік, ізоляцію, безпеку та контроль пристрою.

Видимість та аналітика/аналіз (*Visibility and Analytics*). Не можна боротися з загрозою, яка невидима і яку ви не розумієте. Тому сьогодні пропонуються такі інструменти, як традиційне керування інформацією про безпеку (*security information management* – SIM – це популярний вираз, що означає метод збору, організації та складання звітів із записів, пов'язаних з інформаційною безпекою [29]), а також більш просунуті платформи аналізу безпеки, системи аналізу поведінки користувачів з погляду безпеки (*security user behavior analytics – SuBa*) та інші аналітичні системи, які дозволяють фахівцям з безпеки знати та розуміти, що відбувається у мережі. Ці інструменти, платформи, системи допомагають аналітику з безпеки ретельно стежити за існуючими загрозами та грамотніше організовувати захист. Хоча сьогодні немає жодної платформи, яка охоплювала б необхідну широту функціональності. Але ця область сьогодні активно розвивається [6]. Таким чином, *видимість та аналітика/аналіз* усередині ZTX – це використання та подання даних по всьому підприємству для підтримки обґрунтованих рішень щодо безпеки на основі контекстної інформації.

Автоматизація (*automation*) та *оркестрування/оркестрація* (*orchestration*) у ZTX необхідні для автоматизації ручних процесів та їх зв'язку з політикою безпеки та діями з реагування. *Оркестрування* – це: а) автоматизоване конфігурування/налаштування, координація та керування комп'ютерними системами та програмним забезпеченням [30]; б) шаблон взаємодії, якому повинен слідувати агент веб-служби для досягнення своєї мети; *оркестрування* визначає послідовність та умови, в яких одна веб-служба викликає інші веб-служби для реалізації деякої корисної функції [31]). Вважається, що елемент автоматизації та оркестрування має вирішальне значення для успіху парадигми нульової довіри. Нульова довіра за своєю суттю є динамічною та адаптивною, і єдиний спосіб досягти цього – це автоматизувати та оркеструвати все корпоративне середовище. Можливість мати ефективне керування та контроль над багатьма компонентами, що використовуються в рамках стратегії нульової довіри, є життєво важливою частиною ZTX.

Сьогодні Агентство з кібербезпеки та безпеки інфраструктури (CISA – Cybersecurity and Infrastructure Security Agency), а також Адміністративно-бюджетне управління США (OMB – US Office of Management and Budget – Офіс менеджменту та бюджету США) визнають ці сім компонентів і додають ще один: керівництво (governance) [32]. На рис. 2 наведено зіставлення компонентів ZTX Forrester із стовпами CISA для сучасної нульової довіри.

На рис. 2 (праворуч) показані такі компоненти (стовпи CISA), як *ідентичність*, *пристрої, мережі*, *програми* (*застосунки*) та *робочі навантаження*, а також *дані*. Компоненти «*Видимість та аналітика/аналіз*», «*Автоматизація та оркестрування*» та «*Керівництво*» надають можливості для інтеграції передових досягнень по кожному з наведених вище п'яти компонентів. Розглянемо їх докладніше (у тому числі, уточнюючи раніше дані деяким з них визначення та характеристики).

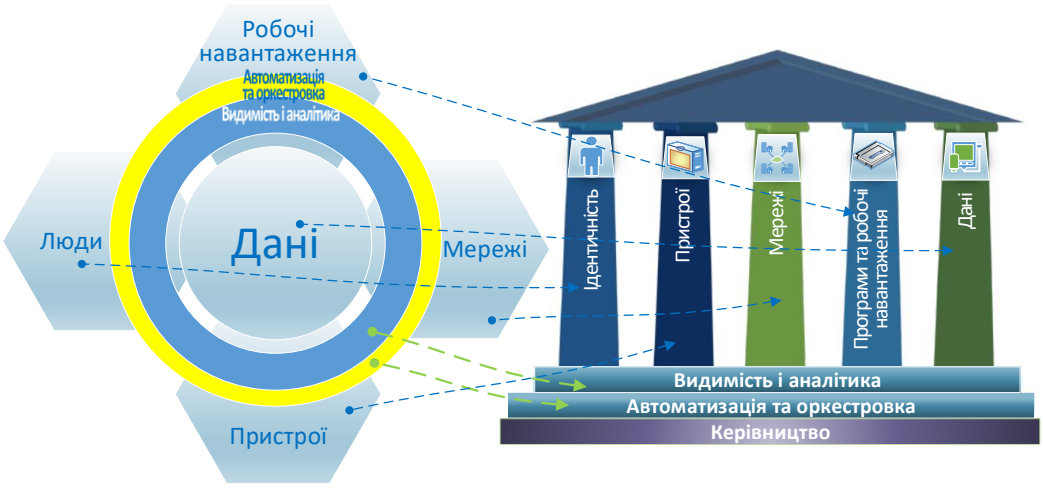


Рис. 2. Зіставлення розширеної моделі нульової довіри Forrester зі стовпами CISA

Ідентичність (*identity*) – це: а) атрибут або набір атрибутів, які однозначно описують користувача або сутність (суб'єкт; під сутностями розуміються користувачі, служби, дані, комп'ютери тощо [33]) підприємства/організації, включаючи сутності, що не є фізичними особами [27]; б) набір значень атрибутів (характеристик), за якими можна розпізнати сутність та які в рамках відповідальності/повноважень спеціаліста з управління ідентифікацією достатні для того, щоб відрізнити цю сутність від будь-якої іншої [34]; в) атрибут або набір атрибутів, які однозначно описують суб'єкт у даному контексті [35, 36].

Говорячи про ідентичність, слід зазначити, що підприємства (організації, компанії, установи) повинні гарантувати та забезпечувати доступ користувачів та суб'єктів до необхідних ресурсів у потрібний час та для певних цілей, не надаючи надмірного доступу. Крім того, вони повинні інтегрувати рішення з керування ідентифікацією, обліковими даними та доступом, за можливістю в рамках всього підприємства, щоб забезпечити сувору автентифікацію, надавати індивідуальну авторизацію (дозволи, повноваження, права) на основі контексту та оцінювати ризики ідентифікації для користувачів та сутностей / суб'єктів цього підприємства. При необхідності організаціям слід інтегрувати свої системи зберігання та управління ідентифікаційними даними, щоб підвищити обізнаність про ідентифікаційні дані підприємства та пов'язані з ними обов'язки та повноваження [32].

Пристрої. Під пристроєм розуміється будь-який актив (включаючи його апаратне та програмне забезпечення, вбудоване мікропрограмне забезпечення (firmware) тощо), який може підключатися до мережі, у тому числі сервери, настільні комп'ютери (desktop) та ноутбуки (laptop), принтери, мобільні телефони, пристрої IoT (internet of things), мережеве обладнання та багато іншого. Пристрої можуть належати установі або бути власністю співробітників (BYOD – bring-your-own-device), партнерів чи відвідувачів. Підприємства повинні забезпечувати безпеку всіх пристроїв підприємства, керувати ризиками авторизованих пристроїв, які не контролюються підприємством, та запобігати доступу неавторизованих пристроїв до ресурсів. Керування пристроями включає ведення динамічної реєстрації всіх активів, включаючи їх апаратне і програмне забезпечення, вбудоване мікропрограмне забезпечення і т. д., а також їх конфігурацій і пов'язаних з ними вразливостей в міру їх виявлення.

Мережі. Під мережею розуміється відкрите комунікаційне середовище, що включає такі типові канали, як внутрішні мережі підприємства (організації, компанії, установи), бездротові мережі та Інтернет, а також інші потенційні канали, такі як стільниковий зв'язок та канали рівня застосунків, що використовуються для передачі повідомлень.

Програми та *робочі навантаження*. Програми (застосунки) та робочі навантаження включають корпоративні системи, комп'ютерні програми та сервіси, які виконуються в локальному середовищі (локально – on-premises), на мобільних пристроях і в хмарних середовищах.

Дані. Дані включають усі структуровані та неструктуровані файли та фрагменти, які знаходяться або знаходились у корпоративних системах, пристроях, мережах, застосунках, базах даних, об'єктах інфраструктури та резервних копіях (включаючи локальні та віртуальні середовища), а також пов'язані з ними метадані.

Компонент «*Видимість та аналітика/аналіз*» забезпечують всебічну видимість, яка є основою для прийняття стратегічних рішень та полегшує дії у відповідь (дії з реагування). Компонент «*Автоматизація та оркестровка*» використовують цю інформацію для підтримки надійних та оптимізованих операцій з обробки інцидентів безпеки та реагування на події у міру їх виникнення. Компонент «*Керівництво*» дозволяє організаціям приймати рішення на основі ризиків. Функції «*Керівництва*» пов'язані із забезпеченням організацій відповідними фахівцями, необхідними технологіями для виконання поставлених завдань з урахуванням дотримання встановлених вимог та наявних ризиків.

Важливо відзначити, що точка зору CISA спирається на початкову концепцію нульової довіри, запропоновану компанією Forrester [37].

2.2. Модель безперервної адаптивної оцінки ризиків та довіри Gartner

Gartner, Inc. підійшла до нульової довіри через свою модель безперервної адаптивної оцінки ризиків та довіри (CARTA). Метою CARTA є забезпечення безперервної оцінки ризиків, що стосуються користувачів, пристроїв, застосунків, даних та робочих навантажень з точки зору прогнозування (predict), запобігання (prevent), виявлення (detect) та реагування (respond). CARTA була представлена компанією Gartner як розвиток адаптивної архітектури безпеки (рис. 3) [38].

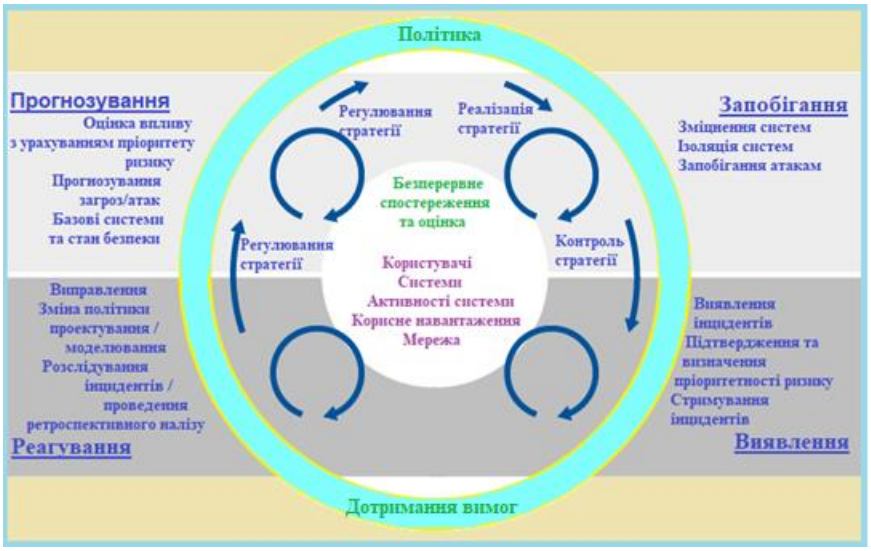


Рис. 3. Чотири складові адаптивної архітектури безпеки

CARTA використовує фундаментальний процес впровадження стратегії безпеки, моніторингу стану/положення (posture) та регулювання/коригування стратегії забезпечення безпеки через різні рівні/площини (planes) безпеки. Gartner вважає, що ці принципи повинні застосовуватися в масштабах всього підприємства і включати вимоги безпеки, політики та відповідності нормативним вимогам.

Gartner прагне розглядати нульову довіру більш вузько, використовуючи терміни доступу до мережі з нульовою довірою (Zero Trust Network Access – ZTNA) для забезпечення безпеки між користувачем та сервером та сегментація мережі з нульовою довірою (Zero Trust Network Segmentation – ZTNS) для мікросегментації / міжсерверної безпеки [6].

3. Основні принципи нульової довіри, запропоновані відомими міжнародними організаціями та компаніями

Три основні принципи – три фундаментальні концепції нульової довіри, представлені в роботі [16], опублікованій компанією Forrester, формулюються таким чином:

1. *Забезпечте безпечний доступ до всіх ресурсів, незалежно від їхнього розташування*.

Це ємне, стисло твердження, яке охоплює безліч аспектів, і в першу чергу необхідність включення всіх ресурсів у сферу компетенції рішення з нульовою довірою. Цей принцип вимагає, щоб рішення з нульовою довірою забезпечувало безпечний доступ усіх облікових записів (людини та комп'ютера) до всіх ресурсів (даних, застосунків, серверів) незалежно від місцезнаходження ідентифікатора (облікового запису сутності) та незалежно від місцезнаходження або використовуваної технології ресурсу, до якого здійснюється доступ виходячи з припущення, що весь трафік є трафіком загрози (принаймні доти, доки немає повної впевненості, що трафік авторизований, перевірений і захищений). У реальних ситуаціях це часто вимагає використання зашифрованих каналів для доступу до даних як у внутрішніх, так і в зовнішніх мережах. Що стосується другої частини принципу – «незалежно від розташування», це стає особливо актуальним в умовах переходу до хмарних технологій, коли більшість даних знаходиться поза традиційними центрами обробки даних. У цьому випадку рішення з нульовою довірою допоможе забезпечити дотримання питань, пов'язаних із розміщенням даних, відповідно до нових правил конфіденційності даних, що з'являються у всьому світі. Саме цей принцип, по суті, вимагає ліквідації традиційного корпоративного периметра та заміни його альтернативною парадигмою безпеки.

2. *Використовуйте стратегію найменших/мінімальних привілеїв та суворо дотримуйтесь принципів контролю доступу*.

Стратегія найменших привілеїв при доступі до ресурсів не є чимось новим, але до появи концепції нульової довіри її важко було реалізувати в широких масштабах. Історично склалося так, що рішення в галузі інформаційної безпеки не могли усунути розрив між забезпеченням безпеки на рівні мережі та застосунків. Сьогодні найменші привілеї слід постійно використовувати в різних місцях і типах ресурсів, а також на мережевому та прикладному рівнях, використовуючи контекст безпеки та ідентифікуючі дані. Однак традиційно користувачі та їх пристрої отримували широкий доступ до мереж (сьогодні вважається, що можливість відправляти мережеві пакети в систему є привілеєм і керувати нею необхідно відповідним чином [6]), а застосунки використовували контроль доступу лише за допомогою перевірки автентичності (автентифікації). Грамотно забезпечуваний контроль доступу (наприклад, використання контролю доступу на основі ролей (RBAC) для всіх співробітників, реалізація керування привілейованими ідентифікаційними даними (PIM – privileged identity management) для доступу до важливих систем тощо) допомагає усунути людську спокусу отримати доступ до обмежених ресурсів. Хоча концепція нульової довіри не визначає RBAC як кращу методологію керування доступом. Сьогодні й в майбутньому слід розглядати й інші технології, методології (наприклад, ту саму ABAC) для керування доступом. У цьому випадку важлива сама концепція мінімальних привілеїв та суворого контролю доступу. Також важливо, щоб фахівці з безпеки розробили відповідну стратегію керування ідентифікацією та доступом, щоб періодично переглядати та підтверджувати права доступу працівників, у тому числі з великими привілеями. Так, наприклад, співробітники, які мають адміністративний доступ до важливих застосунків і систем, можуть завдати шкоди компанії, якщо вони мають злий намір. Крім того, такі привілейовані користувачі часто стають мішенню для хакерів, які намагаються скомпрометувати їхні облікові дані з корисливою метою. У цьому випадку рішення PIM, якраз, дозволяють фахівцям з безпеки уважно стежити за діями зазначених користувачів та вимагати від них пред'явлення паролів щоразу для отримання доступу до важливих систем.

3. Перевіряйте та реєструйте весь трафік.

Мережі є досить важливим місцем в IT-інфраструктурі та забезпеченні безпеки, оскільки вони є сполучною ланкою між розподіленими компонентами та їх взаємодією один з одним. Саме з цієї причини цей принцип потребує перевірки та протоколювання мережевого трафіку. Дійсно, постійно перевіряючи, реєструючи та аналізуючи мережевий трафік, фахівці з безпеки можуть виявити аномальну поведінку користувачів або їхню підозрілу активність (наприклад, якщо користувач виконує великі завантаження або часто звертається до систем або записів, які зазвичай йому не потрібні для виконання повсякденних обов'язків). Технологія мережі з нульовою довірою спрощує передачу вмісту мережевого трафіку та журналів до інструментарію аналізу безпеки для більш глибокого дослідження. Інформація про мережевий трафік повинна бути доповнена системою нульової довіри (додані ідентифікаційні дані та відомості про пристрій) та передана у міжмережеві екрани нового покоління, засоби мережевого моніторингу та SIEM (керування інформацією та подіями безпеки), щоб підвищити їхню здатність приймати рішення щодо виявлення, оповіщення та реагування. Важливо відзначити, що системи нульової довіри повинні не тільки широко вивчати та реєструвати метадані мережевого трафіку, але бути більш уважними при аналізі вмісту мережного трафіку через витрати на обробку та зберігання.

Дані принципи, на думку фахівців у галузі інформаційної безпеки, зазвичай вважаються основними і важливими [6] і повинні дотримуватися в будь-якій реалізації концепції нульової довіри.

Звичайно, як уже зазначалося, значний вплив на галузь інформаційної безпеки в цілому, і концепцію нульової довіри зокрема, зробив вихід у 2020 р. публікації NIST про архітектуру нульової довіри [7] та пов'язаний з нею проєкт NCCoE [4].

З точки зору NIST, *нульова довіра* являє собою набір концепцій та ідей, розроблених для мінімізації невизначеності в застосуванні точних рішень щодо доступу з найменшими привілеями для кожного запиту в інформаційних системах і службах, коли мережу вважають скомпрометованою. Архітектура нульової довіри (ZTA) – це план кібербезпеки підприємства, який використовує концепції нульової довіри та охоплює зв’язки компонентів, планування робочого процесу та політики доступу. Архітектура нульової довіри розробляється та розгортається з дотриманням наступних основних принципів нульової довіри [7]:

1. *Усі джерела даних і обчислювальні послуги вважаються ресурсами*. Мережа може складатися з кількох класів пристроїв. Також підприємство може ухвалити рішення вважати пристрої, що належать особисто співробітнику, ресурсами, якщо вони можуть отримати доступ до ресурсів, що належать підприємству.

2. *Усі комунікації захищаються незалежно від розташування мережі*. Розташування в мережі саме по собі не означає довіри. Довіра не повинна надаватися автоматично на основі того, що пристрій знаходиться в мережевій інфраструктурі підприємства. Усі комунікації повинні здійснюватися найбільш безпечним способом, забезпечувати конфіденційність та цілісність, а також автентифікацію джерела.

3. *Доступ до окремих ресурсів підприємства надається на основі кожного сеансу* (*лише на один сеанс*). Довіра до запитувача оцінюється перед наданням доступу. Доступ також має бути надано з найменшими привілеями, необхідними для виконання завдання.

4. *Доступ до ресурсів визначається динамічною політикою*, включаючи спостережуваний стан ідентичності клієнта, програми/сервісу та активу, який запитується, і може включати інші поведінкові атрибути та атрибути навколишнього середовища. Підприємство захищає ресурси, визначаючи, які ресурси вона має, хто є її членами (або здатність автентифікувати користувачів із об’єднаної спільноти) і який доступ до ресурсів потрібен цим членам. Для нульової довіри ідентифікатор клієнта може включати обліковий запис користувача (або ідентифікатор служби) і будь-які пов’язані атрибути, призначені підприємством цьому обліковому запису для автентифікації автоматизованих завдань. Стан активу запиту може включати такі характеристики пристрою, як встановлені версії програмного забезпечення,

мережеве розташування, час/дата запиту, поведінка, що спостерігалася раніше, і встановлені облікові дані. Політика – це набір правил доступу на основі атрибутів, які організація призначає суб'єкту, активу даних або програмі. Атрибути середовища можуть включати такі фактори, як мережеве розташування запитувача, час, повідомлення про активні атаки тощо. Правила доступу до ресурсів і дозволів на дії можуть відрізнятися залежно від чутливості ресурсу/даних. Принципи найменших привілеїв застосовуються для обмеження як видимості, так і доступності.

5. *Підприємство контролює та вимірює цілісність і стан безпеки всіх належних йому та пов'язаних з ним активів.* Жоден актив не є надійним. Підприємство оцінює стан безпеки активу під час оцінки запиту ресурсу. Підприємство має встановити безперервну діагностику та пом'якшення наслідків (CDM) або подібну систему для моніторингу стану пристроїв та застосунків і застосовувати виправлення за потреби.

6. *Усі автентифікації та авторизації ресурсів є динамічними та суворо контролюються перед тим, як доступ буде дозволений*. Це постійний цикл отримання доступу, сканування та оцінки загроз, адаптації та постійної переоцінки довіри у процесі безперервної взаємодії. Очікується, що підприємство, яке впроваджує ZTA, матиме системи *управління ідентифікацією, обліковими даними та доступом* (ICAM – Identity, Credential, and Access Management) і системи управління активами. Це включає використання *багатофакторної автентифікації* (MFA – multifactor authentication) для доступу до деяких або всіх ресурсів підприємства.

7. *Підприємство збирає якомога більше інформації про поточний стан активів, мережевої інфраструктури та комунікацій і використовує її для покращення стану безпеки*. Підприємство має збирати дані про стан безпеки активів, мережевий трафік і запити на доступ, обробляти ці дані та використовувати будь-яку отриману інформацію для покращення створення та застосування політики.

Ці принципи можна пов'язати з основними компонентами розширеної моделі нульової довіри Forrester. Оскільки підходи до нульової довіри NIST і Forrester відіграють важливу роль як керівництва (які поділяють основні принципові концепції нульової довіри за різними категоріями [39]) для забезпечення безпеки у відповідності до принципу нульової довіри на підприємстві.

У табл. 3 показано зіставлення основних компонент розширеної моделі нульової довіри Forrester з основними принципами нульової довіри NIST, що демонструє їх взаємозв'язок і дозволяє припустити, що будь-який з цих підходів може бути використаний як керівництво для забезпечення безпеки у відповідність до принципу нульової довіри.

Таблиця 3

Порівняння традиційної моделі безпеки та моделі нульової довіри

Forrester	**NIST**						
	Ресурси	Комунікаційна безпека	Безпека сеансу	Контроль доступу	Контроль безпеки активів	Безперервна автентифікація	Реєстрація інформації
Дані		√					
Мережа	√			√			
Люди	√	√	√	√		√	
Робочі навантаження	√			√		√	
Пристрої	√	√	√	√	√	√	
Видимість та аналітика							√
Автоматизація та оркестрування				√		√	

З основними принципами нульової довіри NIST також корелюють вісім принципів нульової довіри Національного центру кібербезпеки Великобританії (NCSC – National Cyber

Security Centre), які також можуть допомогти реалізувати власну архітектуру мережі з нульовою довірою у корпоративному середовищі, і які полягають у наступному [40]:

1. *Вивчіть свою архітектуру, включаючи користувачів, пристрої, послуги та дані*. Знання про кожен компонент своєї архітектури дозволить вам визначити, де знаходяться ваші ключові ресурси, які основні ризики для вашої архітектури, крім того, дозволить уникнути будь-яких помилок на пізньому етапі інтеграції успадкованих / застарілих сервісів, які не підтримують концепцію нульової довіри.

2. *З'ясуйте ідентифікаційні дані ваших користувачів, служб та пристроїв*. Ідентифікаційні дані (identity) можуть представляти користувача (людину), службу (процес) або пристрій. Кожен із них має бути однозначно ідентифікованим в архітектурі з нульовою довірою. Це один із найбільш важливих факторів при ухваленні рішення про те, чи слід комусь або чомусь надати доступ до даних або послуг.

3. *Оцініть поведінку ваших користувачів, стан пристроїв та сервісів*. Поведінка користувачів, стан служб або пристроїв є важливими показниками при забезпеченні впевненості в безпеці ваших систем. Можливість оцінювати поведінку користувачів, працездатність пристроїв та сервісів є ключовим аспектом в архітектурі нульової довіри.

4. *Використовуйте політики для авторизації запитів*. Кожен запит на отримання даних або послуг повинен бути авторизований відповідно до політики. Політики можуть допомогти полегшити управління ризиками під час обміну даними або послугами з гостьовими користувачами або партнерськими організаціями. Механізм політик є ключовим компонентом архітектури нульової довіри, який дозволяє забезпечити гнучкий, адаптований контроль доступу до запитуваних ресурсів.

5. *Автентифікація та авторизація всюди*. Рішення щодо автентифікації та авторизації повинні враховувати безліч ознак, таких як розташування пристрою, працездатність пристрою, особистість та статус користувача, щоб оцінити ризик, пов'язаний із запитом доступу.

6. *Зосередьте моніторинг на користувачах, пристроях та сервісах*. Моніторинг цих пристроїв, сервісів та поведінки користувачів допоможе визначити їхній стан. Моніторинг повинен бути пов'язаний з політиками, які були встановлені для забезпечення впевненості у їх правильному налаштуванні.

7. *Не довіряйте жодній мережі, включаючи власну*. Не довіряйте жодній мережі між пристроєм та сервісом (службою), до якого він звертається, включаючи локальну мережу. При передачі даних по мережі для доступу до даних або служб слід використовувати безпечний транспортний протокол, щоб бути впевненим у тому, що трафік захищений під час передачі і менш схильний до загроз. Архітектура нульової довіри змінює спосіб реалізації традиційних засобів захисту користувачів, таких як фільтрація шкідливих веб-сайтів та захист від фішингу, які можуть забезпечуватись різними рішеннями в архітектурі нульової довіри.

8. *Вибирайте сервіси (служби), створені на основі принципу нульової довіри*. Служби можуть не підтримувати нульову довіру і, отже, можуть вимагати додаткових ресурсів для інтеграції та збільшення витрат на підтримку. У таких випадках доцільно розглянути альтернативні продукти та служби, розроблені з урахуванням принципу нульової довіри. Використання продуктів, в яких застосовуються технології, що базуються на стандартах, дозволяє спростити інтеграцію та взаємодію між службами та постачальниками ідентифікаційних даних.

Ці вісім принципів, наприклад, використовували Oracle при розробці хмарної інфраструктури (OCI – Oracle Cloud Infrastructure) для надання клієнтам вбудованих функцій безпеки, що дозволяють швидко та ефективно захистити їх робочі навантаження [41]. Використання запропонованого Oracle рішення в майбутньому зможе допомогти багатьом організаціям, які хочуть бути більш гнучкими під час трансформації свого бізнесу, використовуючи загальнодоступну хмару для надання економічно ефективної інфраструктури, платформ та програмних послуг, впровадити модель безпеки з нульовою довірою до хмар.

Спираючись на важливий досвід міжнародних організацій та компаній в предметній області, що розглядається, можна зробити висновок, що представлений набір принципів має бути основним для реалізації будь-якої концепції нульової довіри. На наше глибоке переконання, концепція нульової довіри – це правильніший та ефективніший підхід до забезпечення безпеки підприємства. Кожен інтегрований у систему нульової довіри IT-компонент, що має необхідний рівень безпеки, підвищує її ефективність, корисність та сферу дії. І навпаки, кожен ізольований (не інтегрований) компонент створює додаткові труднощі, знижує ефективність системи нульової довіри та може перешкоджати забезпеченню безпеки. Зрозуміло, що немає єдино вірного рішення, спрямованого забезпечення нульового довіри. У зв'язку з цим керівники служб безпеки повинні враховувати інфраструктури, пріоритети, навички персоналу, бюджети та терміни при розробці своєї концепції нульової довіри. Через це концепція нульової довіри може здатися досить складною, але насправді масштаби її застосування допомагають значно вдосконалити систему безпеки та архітектуру підприємства.

В цілому ж, рішення інвестувати у нове рішення безпеки – складне та багатогранне питання. Оскільки інвестиції у безпеку важко оцінити кількісно, оскільки вони часто не приносять очевидної віддачі інвестицій. Особливо важливо враховувати вигоди для бізнес-процесів, співробітників та клієнтів, щоб мати можливість ухвалити обґрунтоване рішення [2].

Висновки

1. Щоб усунути недоліки, притаманні традиційній моделі безпеки (безпеки периметра), було запропоновано концепцію нульового довіри, як філософію, підхід і набір керівних принципів. Основна ідея даної концепції полягає в тому, що жодному учаснику інформаційного обміну не можна довіряти, а будь-який доступ до ресурсів організації є потенційною загрозою. Тому кожен доступ має контролюватись і верифікуватися. При цьому може надаватися повний доступ до служби/сервісу або лише до певних функцій або даних, для яких користувач має право. Причому перевірка не повинна виконуватися тільки на основі пароля, вона повинна враховувати безліч факторів і джерел інформації, таких як: пароль користувача, пристрій, час поточного розташування, права доступу тощо. Важливо визначити політику доступу і суворо її дотримуватися (при цьому політики доступу мають бути динамічними).

2. Планування приведення інфраструктури у відповідність до принципів нульової довіри неможливо здійснити частково або в рамках незначного доопрацювання відповідних інформаційних систем. Потрібна реорганізація інформаційної інфраструктури в цілому, а також інтеграція всіх аспектів, що забезпечують безпеку діяльності організації, щоб принципи нульової довіри показали свою ефективність.

3. Концепція нульової довіри продовжує розвиватися в міру того, як постачальники та організації зі стандартизації переглядають та вдосконалюють специфікації та реалізації нульової довіри, визнаючи це фундаментальним зрушенням у підході до інформаційної безпеки, кібербезпеки.

Список літератури:

1. Department of Defense. Global Information Grid Architectural Vision. Vision for a Net-Centric, Service-Oriented DoD Enterprise. Version 1.0 2007. URL: https://acqnotes.com/Attachments/DoD%20GIG%20Architectural%20Vision,%20June%2007.pdf. .

2. Buck C., Olenberger C., Schweizer A., Völter F., Eymann, T. Never trust, always verify: A multivocal literature review on current knowledge and research gaps of zero-trust // Computers & Security. 2021. 110. 102436.

3. Dhanarani A., Evans R., Loumi H., Lowenthal R., Lopes P., Mesaros M., Schaeumer B., Wahl P., Williams A., Zaidi N. Oracle Database Security a technical primer. Fifth edition. Version 5.0. 2023. URL: https://download.oracle.com/database/oracle-database-security-primer.pdf.

4. Kerman A., Borchert O., Rose S., Division E., Tan A. Implementing a zero trust architecture // National Institute of Standards and Technology. 2020. 17 p. URL: https://www.nccoe.nist.gov/sites/default/files/legacy-files/zta-project-description-final.pdf.

5. National Cybersecurity Center of Excellence (NCCoE). Implementing a Zero Trust Architecture. URL: https://www.nccoe.nist.gov/projects/implementing-zero-trust-architecture.

6. Garbis J., Chapman J. W. Zero Trust Security: An Enterprise Guide. Berkeley, CA: Apress, 2021. 300 p.
7. Rose S., Borchert O., Mitchell S., Connelly S. Zero Trust Architecture. NIST Special Publication 800-207. 2020. https://doi.org/10.6028/NIST.SP.800-207.
8. Samaniego M., Deters R. Zero-trust hierarchical management in IoT // 2018 IEEE international congress on Internet of Things (ICIOT). IEEE, 2018. P. 88–95.
9. Ross R., Pillitteri V., Graubart, R., Bodeau D., McQuaid R. Developing Cyber-Resilient Systems: A Systems Security Engineering Approach // NIST Special Publication 800-160. Vol. 2. Revision 1. 2021. 310 p.
10. Zero Trust Security Market Size, Share & Trends Analysis Report By Deployment (Cloud, On-premises), By Security Type (Network, Endpoint), By Authentication, By Organization Size, By Application, By Region, And Segment Forecasts, 2023-2030. Zero Trust Security Market Size & Trends. URL: https://www.grandviewresearch.com/industry-analysis/zero-trust-security-market-report.
11. Grand View Research. Zero Trust Security Market Growth & Trends. URL: https://www.grandviewresearch.com/press-release/global-zero-trust-security-market.
12. Gartner. Press Release. Gartner Predicts 10% of Large Enterprises Will Have a Mature and Measurable Zero-Trust Program in Place by 2026. URL: https://www.gartner.com/en/newsroom/press-releases/2023-01-23-gartner-predicts-10-percent-of-large-enterprises-will-have-a-mature-and-measurable-zero-trust-program-in-place-by-2026.
13. Fortinet. The State of Zero Trust. Report. 2023. URL: https://www.fortinet.com/content/dam/fortinet/assets/reports/report-state-of-zero-trust.pdf.
14. Martinez J. Zero Trust Architecture: 2024 Complete Guide. URL: https://www.strongdm.com/zero-trust.
15. Shore M., Zeadally S., Keshariya A. Zero trust: the what, how, why, and when // Computer. 2021. Vol. 54. № 11. P. 26–35. https://doi.org/10.1109/MC.2021.3090018.
16. Kindervag J., Balaouras S., Mak K., Blackborow J. No More Chewy Centers: The Zero Trust Model Of Information Security. Forrester Research, Inc. 2016. URL: https://crystaltechnologies.com/wp-content/uploads/2017/12/forrester-zero-trust-model-information-security.pdf.
17. Saltzer J. H., Schroeder M. D. The protection of information in computer systems // Proceedings of the IEEE. 1975. 63(9). P. 1278–1308.
18. Jericho Forum Commandments. Version 1.2. 2007. URL: https://collaboration.opengroup.org/jericho/commandments_v1.2.pdf.
19. Cunningham C., Balaouras S., Barringham B., Dostie P. The Zero Trust eXtended (ZTX) Ecosystem. Extending Zero Trust Security Across Your Digital Business. Forrester Research, Inc. Cambridge, MA. 2018. URL: https://www.cisco.com/c/dam/m/en_sg/solutions/security/pdfs/forrester-ztx.pdf.
20. Ward R., Beyer B. Beyondcorp: A new approach to enterprise security // login. 2014. 39(6). P. 6–11.
21. Osborn, B., McWilliams, J., Beyer, B., Saltonstall M. Beyondcorp: Design to deployment at google // login. 2016. 41(1). P. 28–35.
22. Cittadini L., Spear B., Beyer B., Saltonstall M. Beyondcorp: The access proxy // login. 2016. 41(4). P. 28–35.
23. Peck J., Beyer B., Beske, C. M., Saltonstall M. Migrating to BeyondCorp: maintaining productivity while improving security // login. 2017. 42(2). P. 49–55.
24. Escobedo V., Beyer B., Zyzniewski F., Saltonstall, M. BeyondCorp: the user experience // login. 2017. 42(3). P. 38–43.
25. King H., Janosko M., Beyer B., Saltonstall M. Beyondcorp 6: Building a healthy fleet // login. 2018. 43(3). P. 24–30.
26. Gonçalves G., O'Malley K., Beyer, B., Saltonstall M. BeyondCorp and the long tail of Zero Trust // login. 2023. 52423. URL: https://www.usenix.org/publications/loginonline/beyondcorp-and-long-tail-zero-trust.
27. Continuous Adaptive Risk and Trust Assessment (CARTA). URL: https://www.ssh.com/academy/iam/carta.
28. Sarkar S., Choudhary G., Shandilya S. K., Hussain A., Kim H. Security of Zero Trust Networks in Cloud Computing: A Comparative Review // Sustainability. 2022. 14. 11213. https://doi.org/10.3390/su141811213.
29. Bayuk J. L. Stepping Through the InfoSec Program. ISACA. 2007. 238 p.
30. Erl T. Service-oriented architecture: concepts, technology, and design. Pearson Education India, 2005. 760 p.
31. Singhal A., Winograd T., Scarfone K. Guide to Secure Web Services. Recommendations of the National Institute of Standards and Technology. NIST Special Publication 800-95. 2007. 128 p.
32. Cybersecurity and Infrastructure Security Agency. Zero Trust Maturity Model. Version 2.0. 2023. URL: https://www.cisa.gov/sites/default/files/2023-04/zero_trust_maturity_model_v2_508.pdf.
33. Єсін В. І., Вілігура В. В., Сватовський І. І. Забезпечення безпеки у розподілених інформаційних системах: основні аспекти // Радіотехніка. 2023. Вип. 214. С. 32–63. https://doi.org/10.30837/rt.2023.3.214.04.
34. Boyens J., Bartol N., Boyens J., Moorthy R., Paulsen C., Shankles S. A. Notional supply chain risk management practices for federal information systems // US Department of Commerce, National Institute of Standards and Technology. NISTIR 7622. 2012. 99 p.
35. Committee on National Security Systems (CNSS) Glossary. CNSSI No. 4009. 2022. URL: https://www.niap-ccevs.org/Ref/CNSSI_4009.pdf.
36. Temoshok D., Abruzzi C. Developing trust frameworks to support identity federations. US Department of Commerce, National Institute of Standards and Technology. NISTIR 8149. 2018. 34 p.

37. Holmes D., Burn J., Mellen A., Pollard J., Cerrato P., Cser A. OMB's Zero Trust Strategy: Government Gets Good. URL: https://www.forrester.com/blogs/ombs-zero-trust-strategy-government-gets-good/.

38. van der Meulen R. Build adaptive security architecture into your organization. 2017. URL: https://www.gartner.com/smarterwithgartner/build-adaptive-security-architecture-into-your-organization.

39. Syed N. F., Shah S. W., Shaghaghi A., Anwar A., Baig Z., Doss R. Zero Trust Architecture (ZTA): A Comprehensive Survey // IEEE Access. 2022. Vol. 10. P. 57143-57179. doi: 10.1109/ACCESS.2022.3174679.

40. The National Cyber Security Centre. Zero trust architecture design principles. Guidance. Version 1.0. 2021. URL: https://www.ncsc.gov.uk/collection/zero-trust-architecture.

41. Toal P., Gopalan K. Approaching Zero Trust Security with Oracle Cloud Infrastructure. Version 1.2. Whitepaper. Oracle and/or its affiliates. 2022. URL: https://www.oracle.com/a/ocom/docs/whitepaper-zero-trust-security-oci.pdf.

Надійшла до редколегії 02.06.2024

Відомості про авторів:

Єсін Віталій Іванович – д-р техн. наук, професор, Харківський національний університет імені В.Н. Каразіна, професор кафедри безпеки інформаційних систем і технологій, факультет комп'ютерних наук; Україна; e-mail: v.i.yesin@karazin.ua; ORCID: https://orcid.org/0000-0003-1977-7269

Вілігура Владислав Вікторович – Харківський національний університет імені В.Н. Каразіна, викладач кафедри безпеки інформаційних систем і технологій, факультет комп'ютерних наук; Україна; e-mail: viligura93@gmail.com; ORCID: https://orcid.org/0000-0002-1137-2382

Узлов Дмитро Юрійович – канд. техн. наук, Харківський національний університет імені В. Н. Каразіна, в.о. декана факультету комп'ютерних наук; Україна; e-mail: dmytro.uzlov@karazin.ua; ORCID: https://orcid.org/0000-0003-3308-424X

УДК 004.056.5 **DOI:10.30837/rt.2024.2.217.04**

М.С. КАВЕЦЬКИЙ, О.В. СЄВЄРІНОВ, канд. техн. наук., Р.Ю. ГВОЗДЬОВ, А.О. СМІРНОВ, канд. техн. наук.

ВИКОРИСТАННЯ МАШИННОГО НАВЧАННЯ ДЛЯ КЛАСИФІКАЦІЇ АТАК ТИПУ DOS/DDOS

Вступ

Актуальність роботи проявляється у необхідності виявлення та протидії атакам типу DOS/DDOS, що є серйозною загрозою для сучасних інформаційних систем. Аналіз показав, що ці кібератаки призводять до значних економічних збитків та перерв у роботі мережевих сервісів, підкреслюючи важливість розробки ефективних методів їх виявлення [1, 2]. Сьогодні набирає популярність використання машинного навчання для класифікації різного роду атак. В роботі пропонується використовувати метод дерева прийняття рішень для покращення подібних класифікаторів. Дослідження потенціалу дерев прийняття рішень в цьому контексті вказує на перспективність таких моделей для забезпечення безпеки мереж та запобігання інформаційним загрозам у сучасному цифровому середовищі. Необхідність побудови більш точних класифікаторів для виявлення атак типу DOS/DDOS визначається швидкою еволюцією кіберзагроз та їх все більш складним характером. Точні класифікатори, такі як ті, що базуються на деревах прийняття рішень з оптимально підібраними гіперпараметрами, є ключовим елементом в забезпеченні ефективного захисту мережевих ресурсів і даних користувачів від сучасних кіберзагроз.

Дерева прийняття рішень є одним з методів машинного навчання для класифікації. Алгоритм дерев прийняття рішень легко використовувати за допомогою мови програмування Python та наявності відповідних бібліотек. В Python одним з найпопулярніших пакетів для реалізації алгоритмів дерева прийняття рішень є бібліотека scikit-learn. Вона забезпечує простий і зручний інтерфейс для побудови, навчання та оцінки моделей дерев прийняття рішень, що допомогає швидко та ефективно проводити експерименти з виявлення та протидії кібератакам, а також порівнювати отримані результати, використовуючи загальноприйняті метрики.

Вибір методу дерев прийняття рішень був мотивований його здатністю до адаптації до складних класифікаційних задач та потенціалом у вдосконаленні результатів через правильне налаштування параметрів моделі та відбір оптимального датасету. Використання методу дерев прийняття рішень показало значні покращення у точності виявлення атак, а також в повноті класифікації, що підкреслює його перевагу у контексті обраної проблеми. Також відзначається важливість правильного вибору параметрів моделі та датасету для досягнення найкращих результатів у виявленні атак типу DOS/DDOS. Висновки дослідження підкріплені порівняльним аналізом з іншими методами з інших досліджень, що підтверджує перевагу обраного методу в даному контексті.

Метою роботи є проведення порівняльного аналізу методу дерев прийняття рішень для побудови класифікаторів виявлення атак типу DOS/DDOS. В результаті порівняння побудованих класифікаторів було підтверджено поставлену гіпотезу, а також побудована модель, яка має кращі характеристики з усіх досліджуваних робіт.

Загальні відомості про атаки DOS/DDOS

DoS (Denial of Service) та DDoS (Distributed Denial of Service) атаки є одними з найпоширеніших форм кібератак, що спрямовані на виведення з ладу мережевих ресурсів шляхом перевантаження їх запитами. Вони можуть призвести до недоступності сервісів для легітимних користувачів. DoS атака здійснюється з одного джерела і спрямована на перевантаження цільового сервісу або мережі, використовуючи великі обсяги трафіку або експлуатуючи уразливості системи. DDoS атака є розподіленою, що здійснюється з багатьох

джерел (часто ботнетів), координовано спрямованих на ціль, що робить її більш важкою для виявлення та блокування.

Для виявлення атак типу DoS/DDoS машинне навчання використовується для аналізу комплексних мережевих даних з метою автоматичного виявлення аномальної активності, яка може свідчити про потенційні загрози. Основна перевага цих методів полягає у їх здатності розпізнавати як відомі, так і невідомі типи атак, що є критично важливим для протидії кіберзагрозам, що постійно змінюються.

Машинне навчання для виявлення DoS/DDoS атак базується на аналізі різних параметрів мережі, таких як обсяги трафіку, швидкість передачі даних, час між пакетами, типи запитів і відповідей, а також інші характеристики. Моделі навчаються на історичних даних, де вони виявляють закономірності та шаблони нормальної мережевої активності. Після тренування моделі стають здатними автоматично виявляти відхилення від цих норм, що може свідчити про атаку. Наприклад, модель може виявити збільшену кількість запитів від одного джерела або надмірне використання ресурсів, що є типовими ознаками DoS/DDoS атак. Після виявлення аномалій модель може приймати рішення щодо подальших дій, таких як блокування атакуючих IP-адрес або активація додаткових захисних механізмів.

Аналіз наукових робіт

Автори роботи [3] побудували декілька моделей для виявлення DDOS/DOS атак на базі алгоритмів RF, SVM, NB. І хоча їх задача не була досягти дуже гарної точності (точність RF=0.86, NB=0.79, SVM=0.79), вони описали датасет та провели експеримент з вибором найбільш корисних характеристик з нього. Запевняють, що складність алгоритму класифікації залежить від кількості характеристик в датасеті та кількості даних. Проте, в результаті вони використовували датасет CSICIDS2017, який має 84 унікальні характеристики та обрали всього лише 10 з них. В ході виконання цієї роботи з'явилося припущення, що поганий результат точності класифікації в роботі [3] може бути через вибір занадто малої кількості характеристик. Власне, в цій роботі було обрано більшу кількість характеристик та побудовано на їх основі модель класифікації, яка має більшу точність по всім параметрам (accuracy, recall, precision). Ці метрики наведені далі, а числові результати роботи [3] детально описані в розділі з результатами та порівняннями.

В роботі [4] було використано власноруч зібраний датасет. І хоча методи для виявлення веб-атак були майже такіж самі (SVM, KNN, ANN, NB) як і в роботі [3], але наведене дослідження корисне саме їх експериментом. Суть в тому, що вони будували різні моделі на своєму датасеті, в ході експерименту обираючи різні характеристики з цього датасету. Вони прийшли до висновку, що надійність класифікатору можна сказати визначається ще перед тренуванням, коли обирається набір характеристик, які будуть використовуватися під час тренування. Найбільша точність, яка вийшла в них після вибору певних ознак стала 98.3 % на алгоритмі KNN, проте без вибору ознак вона була 95.67 %. Це на 2.63 % точності менше. І хоча вони отримали непогану точність 98.3 %, проте обрані методи для класифікації, а також результат можна вважати не дуже задовільним в порівнянні з отриманим під час використання дерев прийняття рішень в цій роботі, а також іншого, більш великого, датасету. По іншим метрикам типу Sensitivity, Recall також є просадка, що дає більш загальну картину точності моделі.

Дослідження [5] використовує датасет CICIDS2018, який є оновленою версією CICIDS2017 та містить більший спектр атак. Хоча це не грає ролі для даного дослідження, бо з того датасету беруться дані тільки для атак DOS/DDOS, які є в обох версіях датасетів. Автори роботи «підготували» (опрацювали) для тренування датасет (ці етапи також розглядаються в даній роботі), нормалізували його, та натренували багато різних класифікаторів на різних алгоритмах таких як RF, SVM, KNN, DecisionTree (дерева прийняття рішень), XGBoost. В цій роботі зазначається, що всі алгоритми використовували під час тренування параметри за замовчуванням. Проте, для більшої ефективності моделі необхідно змінювати

ці параметри та підбирати більш оптимальні, щоб результат був краще. Тому і хоча в ході експерименту в них найкраще показав себе алгоритм RF (точність 0.9913), проте зазначається, що в ході експерименту з підбором характеристик з датасету, найбільший приріст у точності дав алгоритм дерев прийняття рішень (DecisionTree), який дав найкращий приріст у точності з 0.9554 (для датасету без обирання певних характеристик) до 0.9895 – тобто, приріст у точності після обрання кращих характеристик з датасету склав 0.0341, де алгоритм RF посів друге місце з приростом всього лише 0.0278. Тож, можна зробити припущення, що алгоритм дерев прийняття рішень для подібної задачі може навіть перевершити результати всіх інших досліджень. Також варто зауважити, що автори дослідження [5] використовували під час тренування параметри за замовчуванням, що є певним недоліком даної роботи, який можна покращити та побудувати модель кращої якості по всім параметрам.

Загалом роботи, які використовують алгоритми машинного навчання для виявлення атак типу DOS/DDOS, дають певне поле для вдосконалення в плані якості моделей. Найбільш гарно показали себе моделі з роботи [5], які давали найбільшу точність. Проте, також ця робота стимулює спробувати покращити результати шляхом використання алгоритму дерев прийняття рішень, бо саме цей алгоритм згідно з цією роботою показав найкращий приріст в точності серед всіх інших. Також сама по собі реалізація має свої недоліки під час тренування, які можна прибрати та побудувати більш точну модель класифікатора. Розглянемо алгоритми, які використовувалися в минулих роботах, а також опис алгоритму дерев прийняття рішень, який був досліджений. Наведено переваги з точки зору теорії даного методу серед інших.

Аналіз алгоритмів машинного навчання в порівнянні з деревом рішень

Розглянемо основний узагальнений принцип дії алгоритмів машинного навчання, які використовувалися в роботах [3 – 5], з їх принципом дії, перевагами та недоліками в розрізі з використаним в цій роботі алгоритмом дерев прийняття рішень.

Принцип дії алгоритму Random Forest можна навести так: це ансамбельний метод, який використовує кілька дерев рішень для класифікації або регресії. Кожне дерево обчислює прогноз, і потім за допомогою голосування або середнього значення результатів дерев формується кінцевий прогноз моделі. Порівнюючи зі звичайним деревом прийняття рішень Random Forest використовує більше одночасно побудованих дерев, що зменшує ризик перенавчання, що може виникнути у випадку глибоких дерев рішень. Це дозволяє отримувати більш стабільні та точні результати, особливо коли даних багато і вони мають велику кількість ознак. Однак його складно інтерпретувати, він довго навчається та потребує багато ресурсів.

Наступний метод Support Vector Machine (SVM) – це метод для розділення класів шляхом знаходження гіперплощини, яка максимально відокремлює дані одного класу від іншого. SVM шукає оптимальну гіперплощину, яка максимізує маржу (відстань між найближчими точками обох класів) і мінімізує помилки класифікації. Порівнюючи з деревом рішень SVM використовує математичні методи для знаходження оптимальної гіперплощини, тоді як Decision Tree розділяє дані на основі порогових значень ознак. SVM може дати кращі результати у випадках, коли дані мають складну неоднорідність, і коли кількість ознак більша за кількість вибірок. Однак може бути чутливий до масштабування ознак і об'єму даних, а також вимагає чіткого налаштування параметрів ядра і регуляризації.

Наївний Баєсівський класифікатор Naive Bayes (NB) використовує теорему Баєса для класифікації об'єктів. Він припускає, що всі ознаки незалежні між собою, що робить його особливо ефективним для текстових даних та категоріальних ознак. Порівнюючи з деревом прийняття рішень наївний баєсівський класифікатор працює на основі ймовірності входження даних до класу, тоді як Decision Tree розділяє дані на основі порогових значень ознак. NB може бути менш точним у складних задачах, де ознаки сильно залежать одна від одної.

Алгоритм k-Nearest Neighbors (kNN) – це метод, що класифікує об'єкти на основі їхніх найближчих сусідів. Він шукає k найближчих точок даних до нового прикладу і визначає клас на основі більшості його сусідів. kNN не вимагає попереднього навчання, він просто зберігає набір даних. В порівнянні з Decision Tree, kNN може бути менш ефективним у великих наборах даних з багатьма ознаками через високу обчислювальну складність. З недоліків можна виділити чутливість до шуму та викидів у даних, а також низьку швидкість визначення класу для великої кількості даних.

Також в роботі [4] використовувалися звичайні штучні нейронні мережі (ANN), які можна сказати моделюють людський мозок, використовуючи шари нейронів, які передають сигнали один одному. Вони складаються з вхідних, прихованих і вихідних шарів, де кожен шар містить нейрони з вагами, які змінюються під час навчання. Однак ANN в порівнянні з деревами рішень вимагає більшої кількості даних для навчання і налагодження параметрів.

Gradient Boosting – це метод, який використовує ансамбль слабких моделей, щоб послідовно покращувати прогноз шляхом оптимізації градієнтного спуску. Кожна наступна модель намагається зменшити помилку попередньої моделі. GB покращує результати шляхом комбінації декількох слабких моделей, тоді як Decision Tree працює незалежно від інших моделей. Але з недоліків можна виділити високу чутливість до високо-змінюваних даних, що може призводити до перенавчання. Також алгоритм вимагає налаштування багатьох гіперпараметрів для досягнення оптимальної продуктивності.

Обраний для дослідження в цій роботі алгоритм Decision Tree. Дерево рішень розділяє дані на основі набору правил, що максимізують інформаційний приріст або зменшують ентропію на кожному кроці. Вузли дерева представляють ознаки, а гілки – умови розбиття. Він дозволяє легко інтерпретувати результати, а також добре працює з категоріальними та числовими даними, однак може бути схильним до перенавчання при занадто глибокому дереві, а також не завжди добре підходить для задач з нерегулярною структурою даних. Проте останнє не грає ролі в даній задачі класифікації атак.

Необхідно більш детально розглянути датасет, який використовувався під час дослідження.

Опис та характеристики використаного датасету

У дослідженні було використано датасет CICIDS2017 [6], який є великим набором даних, який використовується для досліджень в області кібербезпеки, зокрема для виявлення аномалій та атак в мережевому трафіку. Цей набір даних створено Canadian Institute for Cybersecurity і містить різноманітні типи мережевих атак, що робить його дуже корисним для навчання та тестування моделей машинного навчання.

CICIDS2017 був створений для того, щоб надати дослідникам зразки мережевого трафіку, які відображають реальні сценарії атаки та нормальної роботи мережі. Цей набір даних служить важливим ресурсом для розробки і тестування методів виявлення кіберзагроз, включаючи виявлення аномалій, інвазивного аналізу і реагування на інциденти.

Як зазначають розробники датасету, для створення CICIDS2017 використовувалось реальне корпоративне мережеве середовище, що включало сервери, маршрутизатори, комутатори та кінцеві пристрої. Дані були зібрані протягом семи днів, включаючи нормальний трафік, що складається з повсякденних дій користувачів, таких як перегляд веб-сторінок та доступ до баз даних, а також атакуючий трафік, що включав DoS, DDoS, brute force, SQL ін'єкції та сканування портів. Для генерації атак використовувались інструменти, такі як hping3 для DoS та DDoS атак і Metasploit для здійснення експлойтів. Всі мережеві пакети були захоплені за допомогою інструментів Wireshark або Tcpdump, а також велись журнали активності. Дані були ретельно анотовані, що включає мітки для нормального трафіку та різних типів атак, і поділені на навчальні та тестові набори. Це середовище забезпечує достатню кількість прикладів для обох видів трафіку і створює репрезентативний набір даних для моделювання реальних загроз. Тому цей датасет є цілком репрезентативним та відображає реальні дані.

В датасеті наявно багато колонок (які будемо називати характеристиками), що відображають різні аспекти мережевого трафіку. Наприклад, в датасеті наявна колонка dst_port, яка представляє порт призначення, flow_duration, що вимірює тривалість потоку в мілісекундах, та tot_fwd_pkts, яка показує загальну кількість пакетів, що були відправлені вперед у певному потоці. Колонка totlen_fwd_pkts відображає загальну довжину цих пакетів у байтах, а flow_byts_s і flow_pkts_s показують середню кількість байтів і пакетів за секунду відповідно. Всі ці колонки разом дозволяють детально аналізувати поведінку мережевого трафіку, що є важливим для виявлення аномалій та атак.

Процес підготовки даних до тренування

Загалом в датасеті CICIDS2017 міститься багато даних з різними атаками сгрупованими по днях, але в рамках дослідження нас цікавить лише DOS/DDOS, тому було обрано файл «Wednesday-workingHours», який містить 692 тисячі записів атак даного типу. В ньому 79 унікальних характеристик (колонок), за якими можна досліджувати дані та будувати класифікатор.

На рис. 1 наведено повний алгоритм роботи з даними, який необхідний для підготовки датасету до тренування.

Рис. 1. Послідовність роботи з даними для підготовки до тренування

Опишемо більш детально послідовність роботи з даними згідно з рис. 1. Спершу дані потрібно зчитати з файлів. Ці дані були у вигляді файлу «Wednesday-workingHours.pcap_ISCX.csv» в форматі «csv», який можна зчитувати за допомогою бібліотек мови Python.

Далі, наступний етап – розмітка даних на класи. У кожному файлі представлено певний клас трафіку (атака або нормальний), тому додається колонка "label" для маркування: 0 означає відсутність атаки, а 1 – її наявність. Це потрібно тому що алгоритми машинного навчання в основному працюють з числовими даними.

Третій етап передбачає видалення невалідних або некорисних даних. Це включає рядки з пропущеними чи неправильними значеннями, наприклад текст у числових колонках. Некорисні дані, такі як IP-адреси, MAC-адреси, порти, протоколи та мітки часу, також видаляються. Крім того, видаляються дані з надто великими значеннями, що можуть спричиняти помилки, наприклад, числа потоку байтів і пакетів за секунду (flow_byts_s, flow_pkts_s).

Четвертий крок – розподіл даних на тренувальну та тестову частини. Модель навчається на тренувальній частині та тестується на даних, які вона раніше не бачила, щоб оцінити її точність. Дані поділяються на X (ознаки) та Y (мітки класу, атака чи нормальний трафік). В даному випадку розподіл був 80/20, де 20 % датасету пішло на тестувальну вибірку. Таким чином для тренування було 553356 даних, а для проведення тестів – 138339 екземплярів.

П’ятий крок – масштабування даних. Оскільки деякі колонки мають великі значення, а інші – малі, необхідно масштабувати дані, щоб модель не надавала перевагу більшим значенням. Для цього використовується MinMaxScaler, який нормалізує дані. Формула, за якої відбувається масштабування даних:

$$\frac{X-Xmin}{Xmax-Xmin} \tag{1}$$

де X – поточний елемент колонки для масштабування; $Xmin$ – мінімальне значення колонки; $Xmax$ – максимальне значення колонки.

Формула (1) застосовується до кожного елемента кожної колонки у наборі даних. Таким чином, можна отримати нормовані значення, навіть якщо одна колонка оперує числами у діапазоні від 1000 до 10000 та інша від 1 до 10, то в результаті нормування даних, будуть дві колонки, які оперують числами від 0 до 1. Такі числа дуже зручні для тренування, та дають гарні результати, тому що модель дивиться на колонки як на рівноправні за важливістю.

Далі розглянемо показники, за якими можна оцінювати якість моделей, а також опишемо процес тренування класифікатору та наведено порівняння з класифікаторами з інших робіт.

Показники оцінки якості моделей

Наведемо опис характеристик для оцінки якості моделей. За цими характеристиками буде також відбуватися порівняння побудованої в цій роботі моделі з класифікаторами з інших робіт.

Confusion Matrix: таблиця, яка показує кількість правильних і неправильних передбачень, розділених на категорії. Вона містить чотири значення: True Positives (TP) – кількість правильно ідентифікованих атак; False Positives (FP) – кількість нормального трафіку, помилково ідентифікованого як атака; True Negatives (TN) – кількість правильно ідентифікованого нормального трафіку; False Negatives (FN) – кількість атак, які залишилися непоміченими.

Accuracy: частка правильних передбачень серед усіх передбачень. Розраховується за формулою

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

Precision: частка правильних передбачень атак серед усіх передбачених атак. Розраховується за формулою

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

Recall (Sensitivity): частка правильно ідентифікованих атак серед усіх реальних атак. Розраховується за формулою

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

F1 Score: гармонійне середнє між Precision та Recall, що враховує як помилки пропуску атак, так і помилки ідентифікації нормального трафіку як атаки. Розраховується формулою

$$F1 = 2 * \frac{Precision*Recall}{Precision+Recall} \tag{5}$$

Далі наведемо опис підходу до тренування моделі дерева прийняття рішень, а також візуально показано як можна інтерпретувати результати моделі.

Тренування класифікатора на основі дерева прийняття рішень

Після збору та обробки даних настає етап тренування моделі, який потребує певного часу та налаштування гіперпараметрів. Наприклад, для дерева прийняття рішень використовувалися такі параметри: random_state для забезпечення відтворюваності результатів, max_depth для обмеження глибини дерева, min_samples_split для визначення мінімальної

кількості вибірок для розділення вузла та max_features для вказівки кількості ознак для вибору найкращого рішення.

Мова програмування Python має популярну бібліотеку sklearn, в якої використовується багато методів і класів для машинного навчання, включаючи клас DecisionTreeClassifier для дерева прийняття рішень. Використовуючи цей клас, можна передати дані для тренування, налаштувати гіперпараметри та запустити процес тренування за допомогою методу fit().

Тренування відбувається досить швидко, бо самі дані не є дуже великими та алгоритм досить швидкий. Для тренування моделі дерева прийняття рішень обрані наступні гіперпараметри: random_state=1, щоб забезпечити відтворюваність результатів і однаковий розподіл випадковості під час перемішування даних; max_depth=150, що встановлює максимальну глибину дерева, обмежуючи його, щоб уникнути перенавчання; та min_samples_split=15, що визначає мінімальну кількість вибірок, необхідних для розділення вузла дерева, забезпечуючи баланс між складністю моделі та її здатністю узагальнювати нові дані.

Результати моделі показують високу ефективність у виявленні DoS/DDoS атак. Зведена таблиця конфузійної матриці демонструє, що модель правильно класифікувала 87,749 випадків, як нормальний трафік (True Negatives), і 50,552 випадків, як атаки (True Positives). Лише 27 нормальних запитів були помилково ідентифіковані як атаки (False Positives), і лише 11 атак залишилися непоміченими (False Negatives). Загальна точність моделі становить 99.97 %, що свідчить про її високу здатність правильно класифікувати трафік. Показник Precision дорівнює 99.95 %, що означає, що майже всі випадки, класифіковані як атаки, дійсно були атаками. Показник Recall складає 99.98 %, що свідчить про здатність моделі виявляти майже всі реальні атаки. Значення F1 score, яке враховує як Precision, так і Recall, становить 99.96 %, підкреслюючи загальну ефективність моделі.

Аналіз та порівняння результатів моделі зі схожими моделями з інших робіт

Проведемо порівняння та інтерпретуємо результати порівняння зі схожими роботами, котрі будували подібний класифікатор DOS/DDOS трафіку. В табл. 1 можна побачити порівняння результатів тренування моделей з досліджень [3 – 5] з результатами тренування моделі на основі дерева прийняття рішень в цій роботі.

Таблиця 1

Порівняння якості створеної та моделей з інших робіт

Дослідження	Датасет/характеристики	Алгоритм	Метрики			
			Accuracy	Precision	Recall	F1 міра
[3]	CSICIDS2017 / 10	RF	0.8680	0.9963	0.8629	не зазначено
		NB	0.7999	0.8603	0.9006	не зазначено
		SVM	0.7988	0.8436	0.9244	не зазначено
[4]	Власний / 6-10	KNN (8)	0.9830	0.9772	0.9773	0.9770
		NB (10)	0.9487	0.9329	0.9205	0.9201
		SVM (10)	0.9215	0.9023	0.9020	0.9021
		ANN (6)	0.9144	0.8811	0.8772	0.8789
[5]	CICIDS2018 / 26	RF	0.9913	0.9843	0.9992	0.9913
		KNN	0.9886	0.9801	0.9982	0.9885
		SVM	0.9689	0.9583	0.9812	0.9685
		XGBoost	0.9894	0.9806	0.9994	0.9894
		DecisionTree	0.9895	0.9847	0.9947	0.9875
Ця робота	CICIDS2017 / 75	DecisionTree	0.9997	0.9994	0.9997	0.9996

Результати роботи [3] не мають значення F1 міри, проте це не зовсім важливо через доволі низьку точність в побудові класифікатору, що робить подальше порівняння не потрібним. Низьку точність з роботи [1] теоретично могло спричинити невдале обрання характеристик з датасету. Проте, ця робота ставила більше собі за мету показати необхідність обирати правильні характеристики, тож вони могли навмисно взяти таку малу кількість.

В табл. 1 наведено відомості з роботи [4], які були взяті з порівняльної таблиці для методу обирання ознак, який автори прозвали «Wrapper» (вони описали його як підхід до відбору

ознак використовує алгоритм навчання для оцінки корисності підмножини ознак). Робота показала вже більш менш непогані результати, де алгоритм KNN показав найкращі результати в усіх метриках. Точність (Accuracy) складає 98.30 %, що свідчить про високу здатність алгоритму правильно класифікувати вхідні дані. Precision та Recall також мають високі значення (97.72 % та 97.73 % відповідно), що вказує на високу точність та повноту класифікації. F1-міра, яка є гармонійним середнім між Precision та Recall, дорівнює 97.70 %, що підкреслює баланс між цими двома метриками.

Дослідження [5] використовує схожий датасет з дослідженням в цій роботі, проте вони вирішили обрати менше характеристик з датасету, але більше ніж в інших роботах [4, 5]. Серед використаних алгоритмів найкращі результати показав алгоритм Random Forest (RF) з Accuracy 0.9913, Precision 0.9843, Recall 0.9992, та F1 міра 0.9913. Також високі показники продемонстрував алгоритм K-Nearest Neighbors (KNN) з Accuracy 0.9886, Precision 0.9801, Recall 0.9982, та F1 міра 0.9885. Інші алгоритми, такі як SVM, ANN, XGBoost і DecisionTree, також мали значні результати, але поступалися RF і KNN у точності та інших метриках.

Проте, робота [5] також показала, що хоча й алгоритм RF набрав більше точність, алгоритм дерев прийняття рішень показував найкращий приріст від зміни характеристик датасету. Цього не відображено в фінальній табл. 1, але як вже було сказано в минулих розділах – алгоритм дерев прийняття рішень дав найкращий приріст у точності з 0.9554 (для датасету без обирання певних характеристик) до 0.9895 – тобто, приріст у точності після обрання кращих характеристик з датасету склав 0.0341, де алгоритм RF посів друге місце з приростом всього лише 0.0278. Також в цій роботі був явний недолік. Автори зазначили, що навчали всі алгоритми з параметрами за замовчуванням, але для більшої якості моделі необхідно постійно обирати параметри, підбираючи найкращі, які дадуть більшу точність. У власній реалізації дерев прийняття рішень так і відбулося, що дало кращі результати.

Найкращий результат отримала власна реалізація з алгоритмом DecisionTree. Точність (accuracy), яка вимірює частку правильних передбачень серед усіх передбачень досягла значення 0.9997, Precision (точність позитивних передбачень) показує, скільки з передбачених атак дійсно були атаками – стало 0.9994, Recall (чутливість), який відображає здатність моделі виявляти всі реальні атаки став 0.9997, а F1 міра є гармонійним середнім між Precision і Recall, що дає збалансовану оцінку якості моделі набула значення 0.9996. У порівнянні з іншими роботами, жоден алгоритм не досяг такого рівня точності, точності позитивних передбачень, чутливості та гармонійного середнього, як DecisionTree у власній роботі. Такі результати можна пояснити вибором іншого датасету, відбором релевантних характеристик з датасету, а також більш вдалим алгоритмом для цієї задачі, яким є дерево прийняття рішень. Також головного «конкурента» з роботи [3] вдалося перевершити завдяки підбиранню правильних параметрів до алгоритму під час навчання. Отже, можна сказати, що поставлена гіпотеза підтвердилася.

Висновки

У дослідженні вивчено ефективність методу дерев прийняття рішень для виявлення атак типу DOS/DDOS. Було проаналізовано роботи [3 – 5], які дали розуміння, що саме дерева прийняття рішень мають кращий потенціал серед інших методів, якщо обирати правильні характеристики з датасету, а також налаштовувати параметри тренування моделі. Обрані в цій роботі параметри для тренування були наступні: random_state=1, щоб забезпечити відтворюваність результатів і однаковий розподіл випадковості під час перемішування даних; max_depth=150, що встановлює максимальну глибину дерева, обмежуючи його, щоб уникнути перенавчання; та min_samples_split=15, що визначає мінімальну кількість вибірок, необхідних для розділення вузла дерева, забезпечуючи баланс між складністю моделі та її здатністю узагальнювати нові дані.

Результати моделі показують високу ефективність у виявленні атак типу DoS/DDoS. Зведена таблиця конфузійної матриці відображає, що модель правильно класифікувала

87,749 випадків як нормальний трафік (True Negatives) і 50,552 випадки як атаки (True Positives). Лише 27 нормальних запитів були помилково ідентифіковані як атаки (False Positives), і лише 11 атак залишилися непоміченими (False Negatives). Загальна точність моделі становить 99.97 %, що підкреслює її високу здатність правильно класифікувати трафік. Precision досягає 99.95 %, що свідчить про те, що майже всі випадки, визнані моделлю як атаки, є дійсно атаками. Recall складає 99.98 %, що підтверджує здатність моделі виявляти практично всі реальні атаки. Значення F1 score, яке узагальнює як Precision, так і Recall, становить 99.96 %, що підкреслює загальну ефективність моделі.

Також побудована модель має приріст по точності 0.0102 (близько 1 %) з 0.9895 до 0.9997 серед досліджуваних дерев прийняття рішень, а також приріст точності 0.0084 порівнюючи з найкращою реалізацією серед аналізованих – приріст з 0.9913 (точність кращого алгоритму KNN з роботи [5]) до 0.9997 (точність побудованої в рамках роботи моделі).

Порівняльний аналіз з результатами інших досліджень підтвердив переваги використання дерев прийняття рішень у виявленні атак DOS/DDOS. Звичайно, за умови вибору правильних параметрів під час навчання (бо від цього залежить наприклад крок навчання моделі, або кількість ітерацій по датасету, та багато іншого, що потрібно підбирати), вибору відповідного датасету (він повинен якогомога краще характеризувати типи атак, щоб під час навчання модель мала більше узагальчующих даних, які описують ту чи іншу атаку), а також його правильної попередньої обробки (видалення з даних викидів, обробка числових колонок та інше). Проведена робота показала, що вибір іншого датасету та відбір релевантних характеристик значно вплинули на досягнення найкращих результатів. Додатково, підбір оптимальних параметрів моделі під час навчання підсилив ефективність нашої системи порівняно з конкурентами.

Використання даного методу в системах безпеки мереж може значно покращити їхню здатність до виявлення та захисту від кібератак.

Список літератури:

1. Сєвєрінов О.В., Шевцов В.О., Сокол-Кутиловська А.С. Аналіз сучасних методів атак на електронні ресурси органів управління // Системи озброєння і військова техніка. 2017. №1. С. 65–68.
2. Сєвєрінов О.В., Хрєнов А.Г., Поляков А.О. Аналіз сучасних методів атак на автоматизовані системи управління військами та інформаційні мережі // Системи обробки інформації. 2015. №9. С.101–104.
3. Amer A. Abdulrahman, Mahmood K. Ibrahem, Evaluation of Ddos Attacks Detection in a CICIDS2017 Dataset Based on Classification Algorithms [Електронний ресурс]. Режим доступу: https://www.academia.edu/71363307/Evaluation_of_Ddos_Attacks_Detection_in_a_CICIDS2017_Dataset_Based_on_Classification_Algorithms
4. Polat H., Polat O., Cetin A. Detecting DDoS Attacks in Software-Defined Networks Through Feature Selection Methods and Machine Learning Models // Sustainability. 2020. 12(3). P.1035. https://doi.org/10.3390/su12031035
5. Liu Z., Wang Y., Feng F., Liu Y., Li Z, Shan Y. A DDoS Detection Method Based on Feature Engineering and Machine Learning in Software-Defined Networks // Sensors. 2023. 23(13). P.6176. https://doi.org/10.3390/s23136176
6. IDS 2017 | Datasets | Research | Canadian Institute for Cybersecurity | UNB [Електронний ресурс]. Режим доступу: https://www.unb.ca/cic/datasets/ids-2017.html.

Надійшла до редколегії 05.06.2024

Відомості про авторів:

Кавецький Максим Сергійович – Харківський національний університет радіоелектроніки, магістр кафедри безпеки інформаційних технологій факультет комп'ютерної інженерії та управління; Україна; e-mail: maksym.kavetskyi@nure.ua; ORCID: https://orcid.org/0009-0008-7419-1029

Сєвєрінов Олександр Васильович – канд. техн. наук, доцент, Харківський національний університет радіоелектроніки, професор кафедри безпеки інформаційних технологій, факультет комп'ютерної інженерії та управління; Україна; e-mail: oleksandr.sievierinov@nure.ua; ORCID: https://orcid.org/0000-0002-6327-6405

Гвоздьов Роман Юрійович – Харківський національний університет радіоелектроніки, аспірант кафедри безпеки інформаційних технологій; Україна; e–mail: roman.hvozdov@nure.ua; ORCID https://orcid.org/0000-0002-5408-943X

Смірнов Антон Олександрович – канд. техн. наук, Харківський національний університет радіоелектроніки, доцент кафедри безпеки інформаційних технологій, факультет комп'ютерної інженерії та управління; Україна; e-mail: anton.smirnov@nure.ua, ORCID: https://orcid.org/0000-0003-4121-3902

УДК 004.056.5 **DOI:10.30837/rt.2024.2.217.05**

К.М. ШУЛІКА, Д.С. БАЛАГУРА, канд. техн. наук, З.М. СИДОРЕНКО

АНАЛІЗ МЕТОДІВ ОБХОДУ СУЧАСНИХ СИСТЕМ ЗАХИСТУ КІНЦЕВИХ ТОЧОК

Вступ

Сьогодні неможливо уявити захист даних без використання комплексних рішень для захисту кінцевих точок інформаційної системи організації: серверів і робочих станцій. Вимоги до подібних рішень включають забезпечення прозорості моніторингу процесів і автоматизований пошук аномалій в системах, а також можливість реагувати на інциденти безпеки для спеціалістів команд кібербезпеки [1, 2]. Метою статті є огляд та аналіз методів обходу комплексних рішень для захисту кінцевих точок (EDR), які широко використовуються зловмисниками. В статті виділяються та описуються визначні риси кожного з методів обходу EDR та наводяться рекомендації з протидії ним.

Endpoint Detection and Response

EDR (Endpoint Detection and Response) є типом кросплатформенного програмного забезпечення, що наразі найчастіше використовується для моніторингу подій, формування та формалізації інцидентів безпеки та реагування на інциденти на кінцевих точках інформаційної системи організації. EDR часто використовуються в SOC (Security Operational Center) для забезпечення безпеки в масштабі інфраструктури, але існує можливість обходу і цих комплексних рішень [1, 3, 4].

Метод обходу Anti-Malware Scan Interface (AMSI)

Першим і одним з найпоширеніших методів є обхід AMSI. AMSI – це структура Microsoft, яка дозволяє стороннім рішенням для захисту від шкідливих програм мати доступ до компонентів і програм Microsoft, таких як PowerShell, механізми сценаріїв, .NET Framework і WMI. EDR використовують цю структуру для сканування файлів, пам’яті та потоків на наявність зловмисного корисного навантаження. Зловмисники можуть використовувати кілька різних методів, щоб обійти AMSI, наприклад «відображення» (reflection), перехоплення COM-сервера та коригування пам’яті [5].

Схему роботи AMSI наведено на рис. 1.

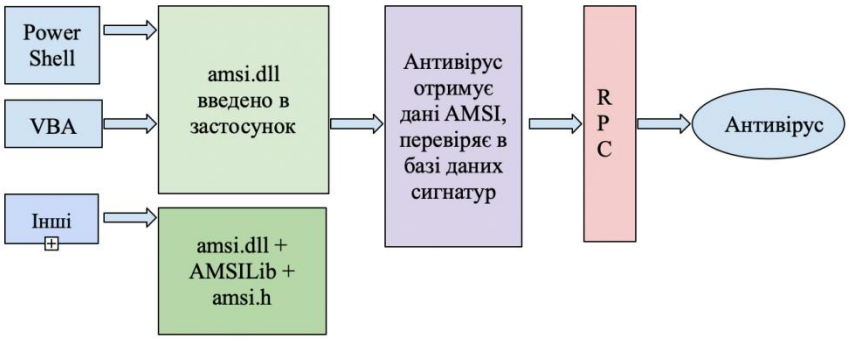


Рис. 1 Схема роботи AMSI

Прикладом нещодавньої атаки з використанням методів обходу AMSI є троян віддаленого доступу Agent Tesla (RAT), який використовував метод коригування пам’яті, щоб уникнути виявлення завантажувача другого етапу атаки та кінцевого корисного навантаження. Перший етап нової версії зловмисного програмного забезпечення включає програму-завантажувач на основі .NET. Завантажувач збирає обфускований код із таких веб-сайтів, як Pastebin і Hastebin. Потім програма встановлення Agent Tesla намагається перезаписати код у

AMSI Microsoft. Спочатку завантажувач намагається отримати адресу пам'яті AmsiScanBuffer (функція Microsoft, також відома як amsi.h, яка сканує буфер на наявність шкідливих програм). Це робиться шляхом виклику amsi.dll Windows за допомоги функції Windows LoadLibraryA, щоб отримати базову адресу DLL. Потім він використовує функцію GetProcAddress для отримання базової адреси та процедуру «AmsiScanBuffer» для отримання адреси функції. Коли Agent Tesla отримує адресу AmsiScanBuffer, він виправляє перші 8 байтів функції в пам'яті. Це змушує AMSI повертати помилку (код 0x80070057), через що всі сканування пам'яті AMSI видаються недійсними. Це блокує програмне забезпечення для захисту кінцевих точок із підтримкою AMSI, фактично змушуючи їх пропускати подальше сканування AMSI для динамічно завантажуваних вузлів у процесі Agent Tesla. Оскільки це відбувається на ранній стадії роботи завантажувача першого етапу, він попереджає виявлення AMSI будь-яких компонентів завантажувача, завантажувача другого етапу та самого корисного навантаження.

В останніх версіях трояну він має додаткову можливість розгортання клієнта Tor для приховування своїх комунікацій, а також використання месенджеру Telegram для викрадання даних. Ці додаткові функції ускладнюють і динамічний, і статичний аналіз антивірусних рішень нового покоління, а також виявлення індикаторів компрометації шкідливого програмного забезпечення (такі як IP-адреси, домени, хеші) для аналітиків з кібербезпеки.

В цьому прикладі варто звернути увагу на те, що механізм обходу EDR може бути вбудований в файли, що мають механізм самозапуску і додавання в автозавантаження при ввімкненні пристрою, тому етап обходу антивірусного ПЗ в системі таким чином перегукується з етапом закріплення на комп'ютері жертви [6, 7].

Зазвичай шкідливе ПЗ Agent Tesla надходить під виглядом вкладення у фішинговий електронний лист. За статистикою Sophos у грудні 2020 р. на Agent Tesla припадало 20 % шкідливих вкладень в електронні листи, а це 1/5 від кількості фішинг-листів, що потрапляють до користувачів Windows. Враховуючи, що Windows наразі залишається найбільш використовуваною операційною системою в світі (69 % всіх користувачів, для порівняння – macOS користується лише 21 % користувачів), а це приблизно 1.4 білліони активних пристроїв, ризик натрапити саме на цей тип трояну із вбудованим обходом AMSI наближається до 1:1 – 45 %.

Запобігання методам обходу AMSI

Методам обходу AMSI можна запобігти, обираючи для корпоративної інфраструктури EDR рішення із вбудованим механізмом захисту, який припиняє процеси, що націлені на компрометацію AMSI, одним із наступних методів: перехопити функцію .NET SetValue, тобто заборонити прямий доступ до змінних amsiInitFailed\amsiSession\amsiContext, або перехопити AmsiUnInitialize() та перехопити PowerShell до та після виконання ScriptBlock, тобто виявити внесення змін в змінну amsiInitFailed (з «False» на «True»), яка не була викликана AmsiUnInitialize(). Про наявність подібного захисту можливо уточнити у постачальника програмного рішення EDR або протестувати самостійно під час тест-драйву.

Також варто розглянути альтернативний шлях: наразі все більше компаній переходять з продуктів Microsoft на інші платформи, такі як macOS та Linux, оскільки вони менш поширені і містять менше вбудованих інструментів, що за своєю архітектурою сприятливі для компрометації [8].

Метод «зняття з крючка» (unhooking)

Наступна техніка, «зняття з крючка» (unhooking) експлуатує той факт, що Windows використовує набір API (наприклад, системний виклик), які можна викликати для виконання інструкцій, що вимагають прямого доступу до системи або рівня ядра. Більшість рішень EDR використовують шлюз ntdll.dll шляхом «зачеплення» до нього, щоб спостерігати за підозрілими зверненнями до пам'яті.

«Зняття з крючка» зловмисники можуть використовувати для завантаження нової непідключеної версії ntdll.dll після того, як Windows завантажила підключену версію в EDR під час запуску процесу. У цей момент EDR не бачить будь-якого коду, який виконується, і не може відстежувати адресу повернення для будь-яких викликів API, що створює так звану «сліпу зону» роботи антивірусного рішення. Якщо зловмисник піде ще далі, то «повторно зачепить» EDR наприкінці своєї операції, щоб приховати факт індикатори своєї присутності [9, 10].

На цьому прикладі зробимо припущення, чому існують методи обходу EDR, майже однакові для рішень від різних вендорів? Тому що EDR використовують однакові шляхи оптимізації роботи агентів і системи в цілому, за рахунок таких рішень забезпечується швидкість їх роботи, але в результаті можуть бути створені подібні «пробоїни».

Запобігання методу «зняття з крючка»

Подібно до прямих системних викликів, метод «зняття з крючка» допомагає зловмиснику уникнути виявлення антивірусом, перешкоджаючи роботі EDR. Цю проблему можна вирішити, запобігаючи модифікації «крючків» (hooks) агента, відстежуючи зміни в цих областях пам'яті. Будь-які спроби отримати доступ до цих областей пам'яті вважатимуться зловмисними та мають блокуватись – про наявність відповідних модулів захисту більше інформації можна отримати від постачальника ПЗ.

Метод завантаження рефлективної DLL

Завантаження рефлективної DLL – це техніка віддаленого запуску коду, використання якої дозволяє зловмиснику завантажити DLL із пам'яті в існуючий процес замість завантаження її з диска. Рішення EDR зазвичай захищають систему, відстежуючи DLL тільки під час їх завантаження з диска, тому завантаження рефлективної DLL забезпечує ще один спосіб уникнення радару EDR. Цей метод часто використовується в поєднанні з одним або декількома іншими техніками та присутній в рамках відомих фреймворків CobaltStrike і Metasploit, які є вагомою частиною інструментарію сучасного хакера. Оскільки обидва фреймворки є надзвичайно поширеними як серед етичних хакерів, так і серед зловмисників, використання інструментів для експлуатації методу завантаження рефлективної DLL є більш ніж доступним за наявності достатніх теоретичних знань.

Запобігання методу завантаження рефлективної DLL

Захиститись від експлуатації цього методу можливо, відстежуючи виділення локальних областей пам'яті, які використовуються під час завантаження рефлективної DLL, і запобігаючи спробам завантаження PE (Portable Executable) файлів із цих областей пам'яті.

Розглянуті в роботі методи обходу EDR та механізми захисту від них представлено в табл. 1.

Таблиця 1

Методи обходу EDR та механізми захисту

Методи обходу EDR	**Механізми захисту**
Обхід Anti-Malware Scan Interface (AMSI)	Захист від перехоплення функції. NET SetValue або перехоплення AmsiUnInitialize() та PowerShell до та після виконання ScriptBlock.
«Зняття з крючка» (unhooking)	Вбудований захист від модифікації «крючків» (hooks) агента.
Завантаження рефлективної DLL	Вбудований захист від виділення локальних областей пам'яті, використовуваних під час завантаження рефлективної DLL, і запобігання спробам завантаження Portable Executable файлів із цих областей пам'яті.

Рекомендації щодо протидії обходу EDR

Якщо постачальник програмного забезпечення не може надати документацію, в якій надано підтвердження стійкості рішення вказаним атакам, то варто розглянути можливість підключення внутрішньої команди компанії для тестування. Узагальнено, тестування EDR рішення на предмет вразливості до описаних типів обходу може бути виконано спеціалістами з тестування на проникнення, а також представниками SOC команди – аналітиками та слідчими з пошуку загроз (threat hunters).

Всі наведені методи можна віднести до стадій атаки Виконання (Execution) та Ухилення від захисту (Defense Evasion) по методології матриці MITRE ATT&CK.

Наведено лише деякі з багатьох способів, за допомогою яких зловмисники можуть обійти захист системи, маючи набір правильних інструментів які легко знайти та застосувати.

Хоча рішення EDR чудово підходять для виявлення відомих зловмисних загроз та їх активності для ідентифікації атаки, коли під час атаки зловмисник намагається викрасти дані та завдати шкоди існує велика прогалина в зупиненні шкідливих процесів та їх підпроцесів, а також в механізмах пошуку перших ознак невідомих загроз, перш ніж вони зможуть потрапити на кінцеву точку. EDR найкраще підходить для виявлення і виконання первинного блокування файлових атак на фізичних комп'ютерах, саме тому не варто розглядати EDR як єдиний інструмент захисту корпоративної мережі.

Роль EDR полягає у виявленні зловмисної поведінки після того, як загроза потрапила у середовище, що надто пізно, щоб зупинити такі складні атаки, як завантаження програми із вбудованим механізмом обходу захисту, тому як загальну рекомендацію можна виділити необхідність доповнювати інструментарій додатковими продуктами для тестування на проникнення, форензики, аналізу мережевого трафіка, зовнішніми та внутрішніми сканерами вразливостей, системами менеджменту корпоративного поштового клієнта, системами автоматизованого реагування на загрози тощо. Також варто звернути увагу на те, що будь-яке антивірусне рішення повинно мати захист від видалення агенту, встановленого на кінцевій точці (tampering protection), що деактивується унікальним токеном. Це дозволить як мінімум отримати не скомпрометовані логи системи, що були передані в консоль EDR, і як максимум – отримати віддалений доступ до скомпрометованої системи для реагування і пошуку слідів зловмисника, без ризику що він видалить агент, маючи права адміністратора пристрою [11].

Профілактика має бути першою лінією захисту, щоб зупинити відомі та виявити ще невідомі загрози, активність програм-вимагачів та загроз нульового дня.

Під профілактикою розуміємо:

1) побудову комплексу з рішень для моніторингу та реагування на інциденти безпеки (XSOAR, в складі якого EDR або XDR, SIEM, IDS/IPS, сканери вразливостей тощо);
2) покращення існуючих процесів, налаштування існуючих інструментів;
3) патч-менеджмент;
4) проведення навчання з кібергігієни для спеціалістів, проведення тестових фішинг-розсилок;
5) ризик-менеджмент та оцінку активності зловмисників в сфері діяльності компанії для попередження невідомих атак;
6) тестування на проникнення і закриття всіх відомих шляхів входу на кінцеві точки компанії;
7) оцінку безпеки партнерів для попередження атаки ланцюга поставок (supply chain).

При дотриманні зазначених рекомендацій можна розраховувати на кращі показники ефективності EDR та покращення загальної стійкості інфраструктури до майбутніх кібератак із використанням методів обходу EDR.

Висновки

Для розуміння наявних ризиків і шляхів їх зниження у статті наведено і проаналізовано три методи обходу EDR що використовуються найчастіше: AMSI обхід, «зняття з крючка»

(unhooking), та завантаження рефлективної DLL. Наведено опис кожного метода, приклади використання в ході атаки на інфраструктуру, а також надані рекомендації щодо протидії та запобіганню використання зловмисниками описаних методів.

Представлені рекомендації можуть бути використані в ході формування процесів в команді з кібербезпеки, для покращення процесів роботи з наявним рішенням EDR, також під час проведення менеджменту ризиків та оцінки профілю інформаційної безпеки організації.

Список літератури:

1. Когут Ю.І. Кібервійна та безпека об'єктів критичної інфраструктури : практ. посіб. Київ : Консалтингова компанія «Сідкон, 2021, С. 132–214.
2. Ушатов В., Сєвєрінов О.В. Проблеми оперативного виявлення і реагування на інциденти інформаційної безпеки. 2019.
3. Sievierinov O., Ovcharenko M., Vlasov A. Enterprise Security Operations Center // Computer and information systems and technologies. 2021.
4. Баклан Я.А., Сєвєрінов О.В. Аналіз систем захисту кінцевих точок від складних загроз EDR // Endpoint Detection and Response. 2022.
5. Antivirus Bypass Techniques: Learn practical techniques and tactics to combat, bypass, and evade antivirus software. Yehoshua Nir, 2021.
6. Malware Tech Blog [Електронний ресурс]. Режим доступу: https://malwaretech.com/2023/12/silly-edr-bypasses-and-where-to-find-them.html.
7. Malware Tech Blog [Електронний ресурс]. Режим доступу: https://malwaretech.com/2023/12/an-introduction-to-bypassing-user-mode-edr-hooks.html.
8. Infosec Write-Ups Blog [Електронний ресурс]. Режим доступу: https://infosecwriteups.com/exploring-antivirus-and-edr-evasion-techniques-step-by-step-part-1-6459563b12ea.
9. IRed Team Blog [Електронний ресурс]. Режим доступу: https://www.ired.team/offensive-security/defense-evasion/bypassing-cylance-and-other-avs-edrs-by-unhooking-windows-apis.
10. Evading EDR: The Definitive Guide to Defeating Endpoint Detection Systems. Matt Hand, 2023.
11. GitHub репозиторій Awesome EDR Bypass [Електронний ресурс]. Режим доступу: https://github.com/tkmru/awesome-edr-bypass.

Надійшла до редколегії 30.05.2024

Відомості про авторів:

Шуліка Катерина Максимівна – Харківський національний університет радіоелектроніки, магістр кафедри безпеки інформаційних технологій; Україна; e–mail: kateryna.shulika@nure.ua

Балагура Дмитро Сергійович – канд. техн. наук, Харківський національний університет радіоелектроніки, доцент кафедри безпеки інформаційних технологій, факультет комп'ютерної інженерії та управління; Україна; e-mail: dmytro.balahura@nure.ua; ORCID: https://orcid.org/0009-0006-9839-3317

Сидоренко Зоя Михайлівна – Харківський національний університет радіоелектроніки, аспірантка кафедри безпеки інформаційних технологій; Україна; e–mail: zoia.sydorenko@nure.ua; ORCID: https://orcid.org/0000-0002-0104-6807

УДК 004.056.5 **DOI:10.30837/rt.2024.2.217.06**

Ю.І. ГОРБЕНКО, канд. техн. наук, Є.В. ОСТРЯНСЬКА

ОЦІНКА ТА ПОРІВНЯННЯ КРИПТОПЕРЕТВОРЕНЬ ТИПУ ЕП НА ОСНОВІ КРИПТОГРАФІЇ НА РЕШІТКАХ КОНКУРСУ NIST США «DIGITAL SIGNATURE SCHEMES»

Вступ

За останнє десятиліття постквантова криптографія досягла переломного моменту; інституційні органи та зацікавлені сторони ініціювали стандартизацію та розгортання, і різноманітні проєкти досягли достатньо високого рівня прогресу та, навіть розгортання та впровадження.

Це підтверджується нещодавньою стандартизацією NIST у 2020 р. геш-підписів XMSS і LMS, а також з цієї причини Національний інститут стандартів і технологій (NIST) проводить конкурс та пропонує перехід до квантово-стійкої криптографії. Протягом трьох раундів було запропоновано багато алгоритмів для шифрування з відкритим ключем, механізмів інкапсуляції ключів і електронного підпису. Для схем електронного підпису було три критерії оцінки: 1) безпека (властивість нульового знання, надійність безпеки в ROM/QROM, спрощення атак бічними каналами, складність основної проблеми), 2) складність і продуктивність та 3) алгоритм і характеристики реалізації на програмнотехнічних засобах.

У липні 2022 р., наприкінці 3-го раунду, щодо постквантових цифрових підписів було запропоновано три кандидати на стандартизацію NIST: один підпис на основі MLWE (Crystals-Dilithium), один підпис на основі NTRU (Falcon) і один підпис на основі гешу (Sphincs+). Хоча профілі ефективності та безпека «чорної скриньки» цих схем добре зрозумілі, стійкість до атак із бічних каналів залишається слабким місцем для всіх них.

Під час атаки бічними каналами криптоаналітик може дізнатися інформацію про фізичне виконання алгоритму, наприклад час його роботи або його вплив на енергоспоживання, а також електромагнітне або акустичне випромінювання пристрою, на якому він запущений. Потім ці допоміжні знання можна використати для відновлення конфіденційної інформації, наприклад криптографічних ключів. Було запропоновано кілька бічних атак проти схем, які NIST розглядає для стандартизації, таких як Dilithium, Falcon або SPHINCS і XMSS. Наведений вище список аж ніяк не є вичерпним, і загалом криптографічні алгоритми вимагають реалізації контрзаходів, щоб досягти будь-якої суттєвої безпеки в контексті атак бічними каналами.

NIST оголосив, що процес стандартизації PQC продовжується четвертим раундом, при цьому наступні KEM все ще знаходяться на розгляді: BIKE, Classic McEliece, HQC і SIKE. Однак на розгляді не залишилося жодного кандидата на цифровий підпис. Таким чином, NIST опублікував заклик до додаткових пропозицій щодо цифрового підпису, які слід розглянути в процесі стандартизації PQC. Прийом документів завершився 1 червня 2023 р.

17 липня 2023 р. NIST оголосив про додаткових кандидатів на цифровий підпис для процесу стандартизації PQC [1].

NIST насамперед зацікавлений у додаткових схемах підписів загального призначення, які не базуються на структурованих решітках. Для певних застосувань, таких як прозорість сертифікатів, NIST також може бути зацікавлений у схемах підписів, які мають короткі підписи та швидку перевірку. NIST відкритий для отримання додаткових матеріалів на основі структурованих решіток, але має намір урізноманітнити стандарти постквантових підписів. Таким чином, будь-яка пропозиція підпису на основі структурованої решітки повинна буде значно перевершувати CRYSTALS-Dilithium [2] і FALCON [3] у відповідних додатках і/або забезпечувати значні додаткові властивості безпеки, які будуть розглянуті для стандартизації.

Метою статті є аналіз, оцінка та порівняння алгоритмів ЕП, в основі яких лежить криптографія на решітках, додаткового конкурсу NIST США. Зокрема, розглянуто алгоритми EagleSign [4], дві версії алгоритму EHTv3 та EHTv4 [5], HAETAE [6], HAWK [7], HuFu [8] та Raccoon [9].

1. Попередні визначення та критерії порівняння

На даний момент NIST обрав 40 алгоритмів-кандидатів ЕП [1]. Серед них є (табл. 1): 6 алгоритмів ЕП на основі кодів, один алгоритм ЕП на основі ізогеній, 7 алгоритмів ЕП, в основі яких лежать операції на решітках, 7 кандидатів на роль алгоритму ЕП на основі методу MPC-in-the-Head та 10 алгоритмів, в основі яких лежать багатоваріативні перетворення, на основі симетричних криптоперетворень було обрано 4 схеми ЕП, та ще 5 кандидатів, що базуються на інших видах криптографічних перетворень.

Таблиця 1

Кандидати на роль криптоперетворення типу ЕП

На основі кодів	На основі ізогеній	На решітках	На основі MPC-in-the-Head	На основі багатоваріативних криптоперетворень	На основі симетричних криптоперетворень	Інше
CROSS	SQIsign	EagleSign	Biscuit	3WISE	AIMer	ALTEQ
Enhanced pqsigRM		EHTv3 та EHTv4	MIRA	DME-Sign	Ascon-Sign	eMLE-Sig 2.0
FuLeeca		HAETAE	MiRitH	HPPC	FAEST	KAZ-SIGN
LESS		HAWK	MQOM	MAYO	SPHINCS-alpha	Preon
MEDS		HuFu	PERK	PROV		Xifrat1-Sign.I
Wave		Raccoon	RYDE	QR-UOV		
		SQUIRRELS	SDitH	SNOVA		
				TUOV		
				UOV		
				VOX		

Фіналістами конкурсу PQC NIST обрав три електронні підписи, серед яких два підписи на решітках. А саме:

- Falcon [3] – це стандарт цифрових підписів PQC (Post Quantum Cryptography), затверджений NIST. Він походить від NTRU і є методами на основі решітки для квантового надійного цифрового підпису. Falcon базується на методі Гентрі, Пейкерта та Вайкунтанатана для створення схем підпису на основі решітки, з використанням швидкої вибірки Фур'є, що дозволяє дуже швидко реалізувати, тисячі підписів за секунду на звичайному комп'ютері; а верифікація відбувається в п'ять-десять разів швидше.

Вибираємо три параметри: N, p і q. Щоб обчислити пару ключів, ми вибираємо два поліноми: f і g. Після цього можемо обчислити $F = f_q = f^{-1}\left(\mathrm{mod}\, q\right)$, де f і f_q – це особисті ключі. Відкритий ключ визначається як $h = p \cdot f_q \cdot f\left(\mathrm{mod}\, q\right)$.

Алгоритм використовує справжню вибірку Гауса, що гарантує незначний витік інформації про секретний ключ до практично нескінченної кількості підписів (більше 2^{64}). Завдяки використанню решіток NTRU підписи значно коротші, ніж у будь-якій схемі підпису на основі решітки з тими самими гарантіями безпеки, тоді як відкриті ключі приблизно однакового розміру. Операції мають вартість $O\left(n \log n\right)$ для ступеня n, що дозволяє використовувати дуже довгострокові параметри безпеки за помірних витрат. Falcon сумісний із невеликими вбудованими пристроями з обмеженим обсягом пам'яті.

- Dilithium [2]. На зараз CRYSTALS (Cryptographic Suite for Algebraic Lattices) підтримує два квантово надійні механізми: Kyber для механізму інкапсуляції ключів (KEM) і обміну ключами; і Dilithium для алгоритму цифрового підпису. CRYSTALS Dilithium використовує схеми Фіата–Шаміра на основі решітки та створює один з найменших підписів серед усіх

постквантових методів із відносно малими розмірами відкритого та закритого ключів. Три основні реалізації параметрів: Dilithium 2, Dilithium 3 і Dilithium 5. Загалом, Dilithium 3 еквівалентний 128-бітному підпису і, можливо, є відправною точкою для реалізації.

2. Алгоритм ЕП EagleSign

Як було сказано вище, NIST стандартизував два методи Dilithium і Falcon для цифрового підпису постквантової криптографії. Dilithium – це підпис на основі MLWE (Module-Learning With Errors), тоді як Falcon використовує підпис на основі NTRU. Тепер є нові методи решітки, які розвиваються як частина раунду додаткового підпису. Одним із них є EagleSign, визначений в [4], який використовує варіацію методу підпису Ель-Гамаля, без переривання, але використовує структуровані решітки.

Більшість відомих методів зламу RLWE та NTRU не можна тривіально узагальнити до відкритого ключа схеми EagleSign. Автори [4] стверджують, що використання разом MNTRU та MLWE в одному відкритому ключі дозволить зробити більш складними алгебраїчні та геометричні властивості основної решітки, і, таким чином, трохи віддалитися від сильних структурованих решіток. У підписі властивість нульового знання гарантує, що процес підписання не розкриває жодної інформації про секретний ключ, пов'язаний із відкритим ключем, який використовується в процесі верифікації.

Підпис є безпечним у ROM [4]. Безпека в ROM випливає із загальної структури, використовуючи лему про розгалуження. EagleSign забезпечує більшу гнучкість для легкого оновлення рівня безпеки в майбутньому

Слід зазначити, що EagleSign швидший і більш простий, ніж Falcon і Dilithium, запропоновані NIST для стандартизації. Для рекомендованих параметрів розміри EagleSign менші, ніж розміри Dilithium, але розміри EaglaSign подібні до розмірів Dilithium для рівнів 2 і 5. Наразі для рівня безпеки 1 алгоритм ще не є реалізованим, як зазначають автори в [4], саме тому в розд. 9 в табл. 5 нижче швидкодія алгоритму наведена лише для 3 та 5-го рівнів безпеки.

Авторами підпису EagleSign [4] доведено безпеку у моделі випадкового оракула, проте не в моделі квантового випадкового оракула. Безпека в моделі випадкового оракула випливає із загальної структури підпису, використовуючи лему про розгалуження. EagleSign забезпечує більшу гнучкість для легкого оновлення рівня безпеки в майбутньому.

3. Алгоритми ЕП EHTv3 та EHTv4

EHT визначає метод цифрових підписів на основі постквантової криптографії. Перша версія алгоритму електронного підпису EHTv1, представлена в [11]. Друга версія EHTv2 з'явилася в матеріалах NISK 2022 [12]. Актуальна версія EHTv3 в основному відрізняється вибором матриці C. Крім того, в [10] представлено EHTv4, який дуже схожий на EHTv3, але використовує арифметику в кільці скінченної групи Z_q замість Z_q. Це забезпечує в цілому більш ефективний алгоритм для порівнянних рівнів безпеки за рахунок більшого розміру підпису. Та якщо порівнювати обидві версії алгоритму EHTv3 та EHTv4 [5] з криптосистемою Dilithium, то в вони мають набагато менші підписи в порівнянні з Dilithium.

Схеми [5] прозорі та прості для розуміння та реалізації. Відкритий ключ генерується за допомогою фактично однієї інверсії матриці та двох множень матриця-матриця. Підпис генерується за допомогою по суті трьох множень матриці-вектору. Верифікація виконується одним множенням матриці-вектору. Оскільки множення матриці легко розпаралелити, реалізація розпаралелювана.

Особистий ключ може бути згенерований із 48-байтового початкового числа, щоб задовольнити рівень безпеки NIST, можна навіть взяти коротше початкове число. У EHTv3 з міркувань ефективності також потрібно зберігати характеристичний поліном (не обов'язково секретний) підматриці C_1 з C. У EHTv4 замість цього можна зберегти інверсію кількох кільцевих елементів групи.

Обидві схеми дозволяють гнучко обирати параметри для підвищення рівня безпеки, якщо це необхідно.

EHTv3 може добре працювати на 8-розрядних платформах, оскільки його арифметика є модулем відносно невеликого додатного цілого числа . Оскільки секретний ключ EHTv3 може бути створений із вихідного коду, здається, що сучасні смарт-карти мають обчислювальні ресурси для реалізації алгоритму генерації підпису EHTv3. Подібне стосується підписів EHTv4. Цей напрямок потребує подальшого вивчення.

4. Алгоритм ЕП HAETAE

HAETAE [6] – це метод постквантової криптографії (PQC) на основі решітки, який базується на методах криптоалгоритму Dilithium, безпека якої базується на складності модульних версій задач LWE та SIS. Таким чином, він використовує підхід «Фіат–Шамір з переривання-ми» [14, 15], що ґрунтується на вибірці відхилення: вибірка відхилення використовується для перетворення пробного підпису, вибірка якого залежить від конфіденційної інформації, у підпис, вибірку якого можна відкрито моделювати. Підпис HAETAE частково схожий на Crystals-Dilithium [2], але відрізняється двома аспектами: бімодальним розподілом для вибірки відхилення (подібно до схеми підпису BLISS [13]) і вибіркою з рівномірного розподілу гіперкулі та відхилення до нього.

Таким чином розміри підписів на 30 – 40 % менші, ніж у Dilithium за порівнянних рівнів безпеки, а ключі перевірки на 20 – 25 % менші. Також неодмінною перевагою є доволі легкі реалізація та впровадження схеми, оскільки весь процес підписання можна реалізувати за допомогою арифметики з фіксованою комою, і значна частина підписання, яка не залежить від повідомлень, може виконуватися «офлайн» для рандомізованої версії схеми.

На відміну від таких алгоритмів як Falcon та Mitaka, що покладаються на цілочисельну вибірку Гауса із довільними центрами, HAETAE має набагато простіший процес генерації ключів, оскільки покладається лише на (з нульовим центром) неперервну вибірку Гауса, що використовується для рівномірної вибірки в гіперкулях. Виклики до нього також можна масово розпаралелювати. Ця відмінність робить HAETAE можливим мати алгоритм підпису з фіксованою комою та легші маскування.

Але необхідно зауважити, що хоча HAETAE простіший з точки зору впровадження, його ключ перевірки та розмір підпису більші, ніж у Falcon, що видно з табл. 2 – 4 відповідно. Та у порівняння з Dilithium алгоритм HAETAE є повільнішим, оскільки алгоритм генерації ключа перезапускається, якщо секретний ключ не задовольняє умову відхилення ключа.

5. Алгоритм ЕП HAWK

HAWK [7] – це метод підпису на основі решітки, який створює підписи за допомогою проблеми ізоморфізму решітки (LIP). Він є швидшим, ніж Dilithium для підписання та верифікації. Він також займає мало пам’яті та підтримується різними апаратними засобами. Наразі немає функції маскування, яка може бути підозрілою для аналізу криптоаналітичних атак бічними каналами. Є деякі занепокоєння щодо доказів безпеки, пов’язаних з HAWK. HAWK схожий на метод FALCON, але використовує іншу складну задачу. Це зводить проблему решітки до задачі найкоротшого вектору, де підпис визначає здатність розв’язувати загальну проблему найближчого вектору. Загалом, однак, він використовує більшу частину коду FALCON для генерації ключів для вирішення рівняння NTRU.

Для генерації ключів у HAWK потрібні вибірки з центрованого біноміального розподілу, який легко отримати з джерела однорідних бітів. Підписування вимагає дискретної гаусової вибірки з фіксованою шириною з двох сумісних класів цілочисельної решітки, що легко досягається за допомогою двох фіксованих попередньо обчислених таблиць достатньої точності.

Схема ЕП HAWK є SUF-CMA безпечною у моделі випадкового оракула за умови складності проблеми omSVP (варіація проблеми пошуку найменшого вектора). Задача відновлення секретного ключа безпосередньо з відкритого ключа є прикладом проблеми ізоморфізму

модульних решіток (smLIP). HAWK має невелике використання пам'яті. Наприклад, HAWK-512 вимагає не більше 14 кілобайтів оперативної пам'яті для будь-якого алгоритму, включаючи більш швидкі варіанти підпису та верифікації. Якщо ключі можуть бути згенеровані ззовні та складно закодовані на пристрої, тоді HAWK-512 і HAWK-1024 можуть підписувати та перевіряти, використовуючи лише 6 КБ і 11 КБ оперативної пам'яті відповідно. Це робить HAWK гарним кандидатом для багатьох вбудованих платформ на основі ядер ARM Cortex-M0(+): продукти в цьому діапазоні включають, наприклад, серію LPC800 від NXP, STM32F0 від ST або XMC1000 від Infineon (16 КіБ SRAM). Окрім цього, будь-яка функція в HAWK, яка залежить від секретних даних, має час роботи, незалежний від цих даних. HAWK добре підходить для різного обладнання, оскільки він не залежить від арифметики з плаваючою комою. Обчислення з плаваючою комою (подвійна точність) не потрібні. Це дозволяє запускати HAWK на багатьох (обмежених) пристроях, не обладнаних таким FPU.

Ефективного маскування для HAWK (поки що) немає. Незважаючи на простоту вимоги лише двох фіксованих дискретних розподілів Гауса для вибірки під час підписання, це відкрита дослідницька проблема для створення ефективного методу маскування на основі таблиці, який використовується. Позитивним моментом є те, що крім цього компонента відомо, як ефективно маскувати решту конструкції HAWK.

6. Алгоритм ЕП HuFu

Існує два основних підходи до підписів на основі решітки: Fiat-Shamir і Hash-and-sign. Загалом, Dilithium використовує підхід Fiat-Shamir, тоді як Falcon використовує підхід Hash-and-sign, використовуючи структуру на решітках GPV. Новим запропонованим стандартом є HuFu.

HuFu [8] – це схема електронного підпису, безпека якої базується на складності стандартних найгірших проблем на загальних решітках. Крім того, що HuFu не використовує структурованих решіток, він має досить інакший дизайн порівняно з Crystals-Dilithium [2] і Falcon [38]. На високому рівні HuFu – це схема підпису типу «геш-і-підпис», запропонована Гентрі, Пейкертом і Вайкунтатаном [18]. Його екземпляр створюється на складних випадкових решітках відповідно до конструкції входу гаджета [19] і використовуючи техніку компактного гаджета [20] для досягнення загальної хорошої продуктивності. У двох словах складові HuFu можна описати так:

HuFu = фреймворк GPV + складні випадкові решітки + компактний гаджет.

Як і у випадку з Falcon, він використовує метод геш-підпису з GPV. Falcon використовує решітку NTRU, тоді як HuFu використовує гаджетний підхід для представлення решітки. За допомогою гаджетного підходу вхід створюється за допомогою лінійного відношення між загальнодоступною решіткою та решіткою гаджета (яка не є повною основою решітки). На жаль, гаджетний підхід призводить до значно більших відкритих і секретних ключів, але може бути основою інших криптографічних примітивів (наприклад, для шифрування на основі ідентифікації та сукупних підписів).

Безпеки алгоритму ЕП HuFu базується на проблемах SIS та LWE, які є принаймні такими ж складними, як стандартні найгірші задачі решітки на загальних решітках. Такі консервативні припущення безпеки уникають ризику алгебраїчних атак на проблеми ідеальної решітки та атаки на підрешітку проти NTRU.

HuFu створено в рамках моделі Micciancio-Peikert [19]. Як результат, HuFu має онлайн/офлайн структуру, а його онлайн-операції прості, швидкі та повністю над цілими числами. Ця функція буде дуже корисною для певних випадків використання. Крім того, структура гаджета забезпечує потужну універсальність, що веде до широкого спектру вдосконалених криптосистем, таких як шифрування на основі атрибутів, підписи груп, сліпі підписи, тощо. Завдяки цьому HuFu легше адаптувати для надання розширеної функціональності.

7. Алгоритм ЕП Raccoon

Raccoon [9] – це схема постквантового підпису на основі решітки, яка використовує метод Fiat Shamir без переривань (на відміну від методу Dilithium, який виконує перетворення Фіат–Шаміра із перериваннями). Цей метод дозволяє підтримувати розподілені порогові підписи [21], а також забезпечує покращену підтримку атак на бічних каналах. Raccoon був розроблений PQShield і був представлений на конкурс NIST PQC для отримання додаткових підписів.

Оскільки структура Raccoon дуже схожа на Dilithium, можна використовувати дуже схожі стратегії реалізації та оптимізації. Особливо, коли використовується 32-розрядна арифметична інструкція «CRT», код NTT для Raccoon по суті еквівалентний коду Dilithium як на мікроконтролерах, так і на SIMD високого класу.

Основний принцип дизайну Raccoon – піддатливість до маскування. По суті, Raccoon можна замаскувати в порядку $d-1$ за допомогою $O(d\log d)$. Це дозволяє маскувати Raccoon на високих рівнях з невеликим впливом на ефективність.

При високих порядках маскування споживання пам'яті стає новим вузьким місцем ефективності через необхідність зберігати поліноми, замасковані у високих порядках. Автори [9] вирішують це за допомогою методів, які дозволяють значно зменшити вартість пам'яті маскованих значень.

Raccoon покладається на (варіанти) припущень на решітках, які добре зрозумілі. А саме Module-LWE та Module-SIS, подібно до (вибраного) основного стандарту Dilithium. Зауважте, що для евклідової норми алгоритм покладається на Module-SIS, на відміну від трохи менш звичайної норми нескінченності, яка використовується в Dilithium.

Ще однією з переваг є проста і портативна реалізація., що є двома основними ідеями дизайну Raccoon. Наприклад, розподіли помилок базуються на рівномірних розподілах по $\{0,\ldots,2^u-1\}$; це робить впровадження простим на широкому спектрі платформ. Подібним чином 49-бітний модуль можна розділити на два 24-бітний і 25-бітний модулі; це полегшує впровадження на 32-розрядних архітектурах.

На відміну від багатьох інших схем, Raccoon не потребує замаскованих реалізацій симетричних криптографічних компонентів, таких як SHA-3/SHAKE. Кількість окремих гаджетів маскування є відносно невеликою, що призводить до простішого та легшого для перевірки мікропрограмного та апаратного забезпечення.

Окрім масштабованості безпеки та теоретичної обґрунтованості, важливою перевагою маскувальних контрзаходів є те, що вони менш залежать від фізичних деталей реалізації порівняно з методами логічного рівня. Таким чином, реалізації – певною мірою – портативні.

Однак алгоритм ЕП Raccoon має більші розміри, ніж Falcon і Dilithium. Завдяки видаленню вибірки відхилень, розмір підпису Raccoon значно більший, ніж Dilithium, незважаючи на те, що він має подібну структуру та базується на подібних припущеннях. Розміри ключів перевірки подібні до розмірів Dilithium. Тобто, якщо порівнювати підпис з Dilithium, то хоча відкритий (перевіряючий) ключ має подібний розмір, особистий (підписуючий) ключ і розмір підпису приблизно в п'ять разів більші у Raccoon.

Це збільшення розміру пов'язане з тим, що розміри підписів Raccoon масштабуються логарифмічно залежно від кількості запитів. На даний момент набори параметрів і відповідні перевірки безпеки для NIST рівнів I, III і V охоплюють максимальну кількість запитів Q_s, що дорівнює 2^{53}, 2^{51} і 2^{55} відповідно.

Ще одним недоліком схеми ЕП Raccoon є те, що вона немає стійкості до атак помилками. Незважаючи на те, що дизайн Raccoon робить його більш стійким до атак із сторонніх каналів, атаки помилками також є серйозною загрозою в реальному конкурентному середовищі.

8. Порівняння алгоритмів ЕП на решітках

У цьому розділі у табл. 2, 3 та 4 наведено результати порівняння розмірів підписів, відкритих та особистих ключів відповідно, вже стандартизованих NIST алгоритмів ЕП на решітках Crystals Dilithium та Falcon з розглянутими вище алгоритмами в додаткового конкурсу NIST. А також обчислення швидкодії розглянутих у попередніх розділах алгоритмів ЕП у табл. 5. Порівняння було проведено для 1, 3 та 5-го рівнів безпеки відповідно. Усі значення в табл. 2 – 4 наведено в байтах, а швидкодія в табл. 5 у циклах процесору.

Таблиця 2

Порівняння розмірів підписів алгоритмів ЕП
для 1, 3 та 5 рівнів безпеки

Схема підпису ЕП	Розмір підпису (у байтах)		
	Рівень безпеки 1	Рівень безпеки 3	Рівень безпеки 5
Crystals Dilithium	2420	3293	4595
Falcon	690	–	1330
EagleSign	2144	2336	3488
EHTv3	169	255	344
EHTv4	369	–	875
HAETAE	1463	2337	2908
HAWK	555	–	1221
HuFu (байт)	2495	3580	4560
Raccoon	11524	14544	20330

Таблиця 3

Порівняння розмірів відкритих ключів алгоритмів ЕП
для 1, 3 та 5 рівнів безпеки

Схема підпису ЕП	Розмір відкритого ключа (у байтах)		
	Рівень безпеки 1	Рівень безпеки 3	Рівень безпеки 5
Crystals Dilithium	1312	1952	2592
Falcon	897	–	1793
EagleSign	1824	1824	3616
EHTv3	83,490	191,574	348,975
EHTv4	1107	–	2623
HAETAE	992	1472	2080
HAWK	1024	–	2440
HuFu (байти)	1083424	2228256	3657888
Raccoon (байтів)	2256	3160	4064

Таблиця 4

Порівняння розмірів секретних ключів алгоритмів ЕП
для 1, 3 та 5 рівнів безпеки

Схема підпису ЕП	Розмір секретного ключа (у байтах)		
	Рівень безпеки 1	Рівень безпеки 3	Рівень безпеки 5
Crystals Dilithium	2528	4000	4864
Falcon	1281	–	2305
EagleSign	573	573	1600
EHTv3	368	532	701
EHTv4	419	–	925
HAETAE	1408	2112	2752
HAWK	184	–	360
HuFu (байти)	11417440	23172960	37418720
Raccoon (байти)	14800	18840	26016

Порівняння швидкодії підписів. Усі запуски алгоритмів та середні оцінки часу було здійснено та розраховано на комп’ютері з 64-розрядною операційною системою Windows 10

на процесорі Intel(R) Core(TM) i7-10510U CPU @ на 2.30 GHz. Дані є усередними над близько 50 запусками кожного алгоритму.

Таблиця 5

Порівняння швидкодії генерації ключів, підпису
та верифікації алгоритмів ЕП для 1, 3 та 5 рівнів безпеки

Схема підпису ЕП	Рівень безпеки	Швидкодія (у циклах)		
		Генерація ключів	Підписання	Верифікація
Crystals Dilithium	1	300,751	1,081,174	327,362
	3	544,232	1,712,783	522,267
	5	819,475	2,383,399	871,609
Falcon	3	19,872,000	396,678	82,339
	5	63,135,000	961,208	205,128
EagleSign	3	1,020,723	1,283,454	955,956
	5	3,443,617	2,358,603	1,602,340
EHTv3	1	465,600,000	181,920,000	1,968,000
	3	1,432,800,000	494,400,000	4,272,000
	5	3,672,000,000	732,000,000	7,584,000
EHTv4	1	29,040,000	21,600,000	9,240,000
	5	276,000,000	142,320,000	62,880,000
HAETAE	1	3,823,188	20,578,698	1,527,304
	3	12,164,364	32,672,248	2,456,500
	5	22,121,374	58,188,178	3,686,662
HAWK	1	8,430,000	85,400	181,000
	5	43,700,000	148,000	303,000
HuFu	1	1,193,896,000	7,322,000	1,804,000
	3	8,916,915,000	18,413,000	6,105,000
	5	9,727,510,000	31,896,000	10,424,000
Raccoon	1	2,256,000	4,817,000	1,757,000
	3	3,252,000	6,860,000	2,764,000
	5	5,199,000	10,062,000	4,554,000

Як видно з таблиць, алгоритми EagleSign та HAWK є доволі швидкими в порівнянні з іншими кандидатами, а HAWK також виграє поміж інших кандидатів для підписання та верифікації. Він також займає мало пам'яті та підтримується різними апаратними засобами. Алгоритм EHTv3 має великі розміри відкритих ключів для всіх рівнів безпеки, але EHTv4 виправляє цей недолік і має розміри ключів та підписів, навіть, менші ніж у Falcon і Dilithium, але в той же час втрачають у швидкодії. Також ми бачимо, що HuFu має найбільші розміри ключів, до чого, на жаль, призводить гаджетний підхід. Але даний алгоритм може бути гарною основою інших криптографічних примітивів (наприклад, для шифрування на основі ідентифікації та сукупних підписів), що потребує подальшого дослідження.

Висновки

1. Розглянуто та проведено порівняння алгоритмів ЕП, в основі яких лежить криптографія на решітках, додаткового конкурсу NIST США [1]. Зокрема, в роботі розглянуто алгоритми EagleSign [4], дві версії алгоритму EHTv3 та EHTv4 [5], HAETAE [6], HAWK [7], HuFu [8] та Raccoon [9]. І в результаті було зроблено наступні висновки.

2. EagleSign простіший і швидший (у деяких випадках), ніж Falcon і Dilithium. Розміри подібні до розмірів Dilithium, але для рекомендованих параметрів розміри EagleSign менші, ніж у Dilithium. Але він має ті самі обмеження, що й будь-який електронний підпис на основі решіток, щодо довгострокової безпеки.

3. Безперечною перевагою EHTv3 і EHTv4 є короткі підписи, створені схемами. Відкритий ключ EHTv4 всього в кілька разів більший за сам підпис. Це явна перевага EHTv4. Але у той же час відкритий ключ EHTv3 досить великий, і це є помітним обмеженням у порівнянні з EHTv4 та деякими іншими схемами підпису на основі решітки.

4. Схема ЕП HAETAE відрізняється від Dilithium двома основними аспектами: використовується бімодальний розподіл для вибірки з відхиленням, як у схемі підпису BLISS, замість «унімодального» розподілу, такого як Dilithium, обираються та відхилюються рівномірні розподіли гіперкуль замість дискретних рівномірних розподілів гіперкуба. Розміри підписів HAETAE на 30 – 40 % менші, ніж у Dilithium за порівнянних рівнів безпеки, а ключі перевірки на 20 – 25 % менші. З точки зору реалізації, незважаючи на те, що обґрунтування її конструкції відрізняється від обґрунтування конструкції Dilithium, схема залишається зручною для впровадження.

5. Схема ЕП HAWK є SUF-CMA безпечною у моделі випадкового оракула за умови складності проблеми omSVP (варіація проблеми пошуку найменшого вектора). Задача відновлення секретного ключа безпосередньо з відкритого ключа є прикладом проблеми ізоморфізму модульних решіток (smLIP). HAWK має невелике використання пам'яті і добре підходить для різного обладнання, оскільки він не залежить від арифметики з плаваючою комою. Обчислення з плаваючою комою (подвійна точність) не потрібні. Це дозволяє запускати HAWK на багатьох (обмежених) пристроях, не обладнаних таким FPU.

6. Алгоритм HuFu має короткі підписи та швидку реалізація: характерний розмір HuFu відповідає розміру Crystals-Dilithium, тоді як HuFu не базується на структурованих решітках. Підписання та верифікація HuFu є ефективними. HuFu також добре розпаралелюється, що дає певний простір для оптимізації. Крім того, його онлайн/офлайн-структура дозволяє ще більше скоротити онлайн-час виконання та обчислювальний ресурс. Але цей алгоритм також має і ряд недоліків. Як видно з табл. 3, 4 HuFu має великі відкриті ключі: розмір відкритого ключа HuFu становить від 1 до 3,5 мегабайт для трьох різних рівнів безпеки, тому HuFu може бути не дуже бажаним для багатьох програм. Тим не менш, HuFu цілком придатний для випадків використання, коли ключі не передаються часто. У той час як онлайн фаза в процедурі підписання реалізована повністю над цілими числами, автономна фаза все ще часто використовує арифметику з плаваючою комою. Це може бути обмеженням, коли режим онлайн/офлайн вимкнено, особливо для реалізацій на пристроях обмежень. Однак це можна вирішити за допомогою техніки інтегрального розкладання за Грамом.

7. Схема підпису Raccoon заснована на перетворенні Фіата–Шаміра з підпису на основі решітки Шнорра з тонким аналізом, щоб запобігти використанню методу відхилення вибірки та вибірки Гауса, дозволяючи використовувати методи маскування для секретних ключів під час алгоритму підпису, як контрзахід проти атак сторонніми каналами. Крім того, безпека Raccoon ґрунтується на стандартних припущеннях щодо модульних решіток, які використовуються в (вибраному) стандартизованому Dilithium. Ці припущення добре вивчені протягом десяти років і показали свою надійність, пропонуючи справедливий розмір параметрів. Одним з основних недоліків схеми ЕП Raccoon є те, що вона немає стійкості до атак помилками. Незважаючи на те, що дизайн Raccoon робить його більш стійким до атак із сторонніх каналів, атаки помилками також є серйозною загрозою в реальному середовищі.

8. Підсумовуючи порівняльний аналіз в розділі 6 можна сказати, що алгоритми EagleSign та HAWK є доволі швидкими в порівнянні з іншими кандидатами, а HAWK також виграє поміж інших кандидатів для підписання та верифікації. Він також займає мало пам'яті та підтримується різними апаратними засобами, що робить його доволі гарним кандидатом. Алгоритм EHTv4 має розміри ключів та підписів, навіть, менші ніж у Falcon і Dilithium, але в той же час втрачають у швидкодії. Також слід зазначити, що HuFu має найбільші розміри ключів, до чого, на жаль, призводить гаджетний підхід. Але даний алгоритм може бути гарною основою інших криптографічних примітивів (наприклад, для шифрування на основі ідентифікації та сукупних підписів), що потребує подальшого дослідження.

9. Алгоритм ЕП Raccoon має більші розміри, ніж Falcon і Dilithium. Завдяки видаленню вибірки відхилень, розмір підпису Raccoon значно більший, ніж Dilithium, незважаючи на те, що він має подібну структуру та базується на подібних припущеннях. Розміри ключів перевірки подібні до розмірів Dilithium. Тобто, якщо порівнювати підпис з Dilithium, то хоча відкритий ключ має подібний розмір, особистий ключ і розмір підпису у Raccoon приблизно в п'ять разів більші. Це збільшення розміру пов'язане з тим, що розміри підписів Raccoon масштабуються логарифмічно залежно від кількості запитів.

Список літератури:

1. NIST standardization process “Post-Quantum Cryptography: Digital Signature Schemes”. Access mode: https://csrc.nist.gov/Projects/pqc-dig-sig/round-1-additional-signatures
2. Léo Ducas, Eike Kiltz, Tancrède Lepoint, Vadim Lyubashevsky, Peter Schwabe, Gregor Seiler and Damien Stehlé. “CRYSTALS-Dilithium: A Lattice-Based Digital Signature Scheme”. 2022.
3. Thomas Prest; Pierre-Alain Fouque; Jeffrey Hoffstein; Paul Kirchner. Falcon: Fast-Fourier Lattice-based Compact Signatures over NTRU. Specification v1.2 – 01/10/2020.
4. Hounkpevi A.C., Djimnaibeye S., Seck M. EagleSign: A new post-quantum ElGamal-like signature over lattices. Submission to the NIST's post-quantum cryptography standardization process. (2023).
5. Semaev I., Feussner M. Digital Signature Algorithms EHTV3 and EHTV4 submission to NIST PQC. Submission to the NIST’s post-quantum cryptography standardization process. 2023
6. Cheon J. H., Choe H., Devevey J., Güneysu T., Hong D., Krausz M., Yi M. Haetae: Shorter lattice-based fiat-shamir signatures // Cryptology ePrint Archive. 2023.
7. Joppe W. Bos, Olivier Bronchain, Léo Ducas, Serge Fehr, Yu-Hsuan Huang, Thomas Pornin, Eamonn W. Postlethwaite, Thomas Prest, Ludo N. Pulles, Wessel van Woerden. HAWK. version 1.0 (June 1, 2023). [Electronic resource]. Access mode: https://hawk-sign.info.
8. Yang Yu, Huiwen Jia, Leibo Li, Delong Ran, Zhiyuan Qiu, Shiduo Zhang, Xiuhan Lin, and Xiaoyun Wang. HuFu: Hash-and-Sign Signatures From Powerful Gadgets. Algorithm Specifications and Supporting Documentation. 2023.
9. Rafael del Pino, Shuichi Katsumata, Thomas Prest, Mélissa Rossi. Raccoon: A Masking-Friendly Signature Proven in the Probing Model. CRYPTO, 2024.
10. A. Becker, N. Gama, A. Joux, Solving shortest and closest vector problems: The decomposition approach, IACR Cryptology ePrint Archive, 2013/685.
11. I. Semaev. New Digital Signature Algorithm EHT, Cryptology ePrint Atchive, 2022/339.
12. I. Semaev. New Digital Signature Algorithm EHTv2, NISK 2022, 28.11-1.12.2022, Kristiansand, Norway.
13. Leo Ducas, Alain Durmus, Tancrede Lepoint, and Vadim Lyubashevsky. Lattice signatures and bimodal Gaussians // Ran Canetti and Juan A. Garay, editors, Advances in Cryptology – CRYPTO, pages 40–56. Springer, 2013.
14. Vadim Lyubashevsky. Fiat-Shamir with aborts: Applications to lattice and factoring-based signatures // Mitsuru Matsui, editor, Advances in Cryptology – ASIACRYPT, pages 598–616. Springer, 2009.
15. Vadim Lyubashevsky. Lattice signatures without trapdoors // David Pointcheval and Thomas Johansson, editors, Advances in Cryptology – EUROCRYPT, pages 738–755. Springer, 2012.
16. Erdem Alkim, Paulo S. L. M. Barreto, Nina Bindel, Juliane Kramer, Patrick Longa, and Jefferson E. Ricardini. The lattice-based digital signature scheme qTESLA // Cryptology ePrint Archive, Number 2019/085, 2019. [Electronic resource]. – Access mode: https://eprint.iacr.org/2019/085.
17. Melissa Azouaoui, Olivier Bronchain, Gaetan Cassiers, Clement Hoffmann, Yulia Kuzovkova, Joost Renes, Markus Schonauer, Tobias Schneider, Francois-Xavier Standaert, and Christine van Vredendaal. Leveling Dilithium against leakage: Revisited sensitivity analysis and improved implementations // Cryptology ePrint Archive, Report 2022/1406, 2022. [Electronic resource]. – Access mode: https://eprint.iacr.org/2022/1406.
18. Craig Gentry, Chris Peikert, and Vinod Vaikuntanathan. Trapdoors for hard lattices and new cryptographic constructions // Richard E. Ladner and Cynthia Dwork, editors, 40th ACM STOC, pages 197–206. ACM Press, May 2008.
19. Daniele Micciancio and Chris Peikert. Trapdoors for lattices: Simpler, tighter, faster, smaller // David Pointcheval and Thomas Johansson, editors, EUROCRYPT 2012, volume 7237 of LNCS, pages 700–718. Springer, Heidelberg, April 2012.
20. Yang Yu, Huiwen Jia, and Xiaoyun Wang. Compact lattice gadget and its applications to hash-and-sign signatures // CRYPTO 2023, page (to appear), 2023.
21. Del Pino R., Katsumata S., Maller M., Mouhartem F., Prest T., & Saarinen M. J. (2024, May). Threshold raccoon: Practical threshold signatures from standard lattice assumptions // Annual International Conference on the Theory and Applications of Cryptographic Techniques (pp. 219–248). Cham: Springer Nature Switzerland.

Надійшла до редколегії 27.05.2024

Відомості про авторів:

Горбенко Юрій Іванович – канд. техн. наук, АТ «Інститут Інформаційних Технологій», перший заступник, головного конструктора, Україна; e-mail: gorbenkou@iit.kharkov.ua, ORCID: https://orcid.org/0000-0003-0073-9107

Острянська Єлизавета Вадимівна – Харківський національний університет імені В.Н. Каразіна, молодший науковий працівник, АТ «Інститут Інформаційних Технологій», аналітик з систем захисту інформації; Україна; e-mail: antelizza@gmail.com

УДК 004.056.55 **DOI:10.30837/rt.2024.2.217.07**

С.О. КАНДІЙ, І.Д. ГОРБЕНКО, д-р техн. наук

ОЦІНКА ВПЛИВУ АЛГЕБРАЇЧНОЇ СТРУКТУРИ Q-АРНИХ РЕШІТОК НА СКЛАДНІСТЬ КРИПТОАНАЛІЗУ ПРОБЛЕМ НА РЕШІТКАХ

Вступ

Криптографія на решітках стала однією з найбільш перспективних областей досліджень у сучасній криптографії [1]. Основною перевагою криптографії на решітках є її стійкість до квантових атак, що робить її особливо актуальною в контексті розвитку квантових технологій. Такі теоретико-числові проблеми як LWE (англ. Learning With Errors) [2], NTRU (англ. N-th Degree Truncated Polynomial Ring) [3] та SIS (англ. Shortest Integer Solution) [4], демонструють високий рівень безпеки і є основою для побудови таких криптографічних примітивів, як механізми інкапсуляції ключів та електронні підписи.

Для оцінки безпеки криптографічних схем на решітках часто використовується модель GSA (англ. Geometric Series Assumption) [5]. Ця модель базується на припущенні, що під час редукції решіток довжини векторів базису в процесі ортогоналізації утворюють геометричну прогресію. Модель GSA дозволяє спрощено оцінити складність розв'язання важливих задач на решітках, таких як LWE та NTRU.

Однак модель GSA є лише грубим наближенням і має свої обмеження [6]. Вона не завжди точно відображає поведінку реальних алгоритмів редукції решіток. Зокрема, GSA не враховує вплив структури q-арних решіток на довжини векторів базису. Тому результати, отримані за допомогою GSA, слід розглядати як орієнтовні оцінки, а не точні значення.

Для підвищення точності оцінки безпеки криптографічних схем на решітках необхідно розробляти більш комплексні моделі. Такі моделі повинні враховувати широкий спектр факторів, включаючи різні підходи до ортогоналізації, вплив випадкових флуктуацій, а також адаптивні стратегії атаки.

Одним з можливих варіантів є використання симуляторів редукції решіток [7 – 9]. Деякими авторами симулятори вже використовувалися для оцінки безпеки [10]. Проте, по-перше, більшість авторів використовували симулятор Чена–Нгуєна [7], який не враховує багато факторів. По-друге, у більшості випадків симулятор використовувався обмежено тільки для оцінки атак вкладення (англ. Primal USVP Attack) [11]. Іноді, наприклад для оцінки ДСТУ 8961:2019 [12], симулятор використовувався для оцінки гібридних атак. Проте, комплексної методики, що враховує вплив алгебраїчної структури q-арних решіток на процеси редукції решіток, досі не було запропоновано.

Ця робота присвячена комплексній оцінці впливу алгебраїчної структури q-арних решіток на складність криптоаналізу проблем LWE, SIS та NTRU для широкого спектру атак на криптографічні схеми на решітках.

1. Попередні відомості

Введемо необхідні позначення з теорії решіток, згідно з [13]. Решітка L з базисом B є множиною цілочисельних комбінацій лінійно незалежних векторів $b_1, \dots, b_n \in \mathbb{R}^n$:

$$L(b_1, \dots, b_n) = \{\sum_{i=1}^{n} x_i b_i | x_i \in \mathbb{Z}\}. \quad (1)$$

Довжиною вектора v є стандартна евклідова норма $\|v\| = \sqrt{v \cdot v}$, де операція $\cdot$ є скалярним добутком і для двох векторів $v = (v_1, \dots, v_n)$ і $w = (w_1, \dots, w_n)$ визначена як $v \cdot w = \sum_{i=1}^{n} v_i w_i$.

Для заданого базису $B = (b_1, \dots, b_n)$ ортогоналізований за Граммом–Шмідтом базис є $B^* = (b_1^*, \dots, b_n^*)$, де $b_i^* = b_i - \sum_{j=1}^{i-1} \mu_{ij} b_j^*$ для $1 \le j < i \le n$, де $\mu_{ij} = (b_i \cdot b_j^*)/\|b_j^*\|^2$ –

коефіцієнти Грамма–Шмідта, $\|b_j^*\|$ – довжини векторів Грамма–Шмідта (ГШ-довжини). Сукупність ГШ-довжин будемо називати профілем базису.

Для решітки $L = L(B) \subseteq \mathbb{Z}^n$ з базисом $B \in \mathbb{Z}^{n\times k}$ фундаментальний паралелепіпед визначений як $P(B) = \{B \cdot x | x \in [0,1)^k\}$. Детермінант базису решітки є інваріантом і може бути обчислений як $\det(L) = \sqrt{\det(B^T B)} = \prod_{i=1}^{n} \|b_i^*\|$. При цьому, детермінант решітки чисельно дорівнює об'єму фундаментального паралелепіпеда $vol(L)$.

Ортогональна проєкція є відображення $\pi_i: \mathbb{R}^n \mapsto span(b_i, \dots, b_{i-1})^{\perp}$ для $i \in \{1, \dots, n\}$. Проективна решітка $L_{[i:j]}$ – решітка, яка задається наступним чином:

$$L_{[i:j]} = B_i = L(\pi_i(b_i), \pi_i(b_{i+1}), \dots, \pi(b_j)), \tag{2}$$

для $j \in \{i, i+1, \dots, n\}$.

У кожній решітці L існує найменший ненульовий вектор. $\lambda_1(L)$ – норма найменшого вектора. Проблема пошуку найменшого вектора (SVP) полягає у пошуку вектора довжини $\lambda_1(L)$.

Проблема LWE. Нехай $n, q > 0$ – цілі числа, χ – деякий розподіл ймовірностей над множиною цілих чисел $\mathbb{Z}$ та s – секретний вектор з рівномірного розподілу над $\mathbb{Z}_q^n$. $L_{s,\chi}$ є розподілом ймовірностей над $\mathbb{Z}_q^n \times \mathbb{Z}_q$, який отримується наступним чином. Обирається вектор $a \in \mathbb{Z}_q^n$ з рівномірного розподілу, значення помилки $e \in \mathbb{Z}_q$ з розподілу χ та повертається пара $(a, c) = (a, (a, a \cdot s + e) \in \mathbb{Z}_q^n \times \mathbb{Z}_q^n$. Проблема LWE (обчислювальна версія) полягає у тому, щоб для поліноміальної кількості пар (a, c) з розподілу $L_{s,\chi}$ знайти вектор s.

Проблема SIS. Нехай задано ціле число $q > 0$, матриця $A \in \mathbb{Z}_q^{n\times m}$, дійсне число B. Необхідно знайти ненульовий вектор $e \in \mathbb{Z}^m$, для якого виконується $A\mathrm{e} = 0\mathrm{mod}\mathrm{q}$ та $\|e\|_2 \leq B$.

Проблема NTRU. Нехай $n, q > 0$ – цілі числа і задано кільце R_q (на практиці – поле, проте у загальному випадку NTRU визначається для довільних кілець) поліномів степеня n над кільцем лишків за модулем q. Нехай $f, g \in R_q$ – поліноми з деякого розподілу χ і $h = g/f$. Проблема NTRU (обчислювальна версія) полягає у пошуку малих поліномів f, g для заданого полінома h.

2. Моделі базису решіток

Надалі вважаємо, що задана деяка (не обов'язково q-арна) d-вимірна решітка Λ з базисом $B = (b_1, \dots, b_d)$ та ГШ-профілем $B^* = (b_1^*, \dots, b_d^*)$.

В основі аналізу сучасних моделей редукції решіток лежить евристика Гауса [13], сутність якої полягає у тому, що кількість $|\Lambda \cap \Omega|$ точок решітки у довільному вимірюваному тілі $\Omega \subset \mathbb{R}^n$ складає $vol(\Omega)/vol(\Lambda)$. Використовуючи d-вимірний шар у якості вимірюваного тіла, для випадкової решітки $\Lambda \subset \mathbb{R}^d$, очікуваний найменший вектор, згідно з евристикою Гауса, можливо оцінити як

$$GH(\Lambda) = \left(\frac{vol(\Lambda)}{vol(\Omega)}\right)^{1/d} = \frac{\Gamma\left(1+\frac{d}{2}\right)}{\sqrt{\pi}} \cdot vol(\Lambda)^{\frac{1}{d}} \approx \sqrt{\frac{d}{2\pi e}} \cdot vol(\Lambda)^{1/d}. \tag{3}$$

Для зручності аналізу також введемо позначення $gh(d) = \sqrt{d/(2\pi e)}$ та $lgh(\Lambda) = \log_2 gh(\Lambda)$.

Практичні експерименти з алгоритмами LLL та BKZ показують [7 – 9], що $\|b_i^*\|/\|b_{i+1}^*\| \approx const$, якщо $d \gg \beta$. Модель GSA використовує це практичне спостереження та

евристику Гауса. Згідно з моделлю GSA, для довільної BKZ-β редукованої решітки з базисом B та об'ємом V має наступну форму:

$$log\|b_i^*\| = \frac{d-1-2i}{2} \cdot \log(\alpha_\beta) + \frac{1}{d} logV, \tag{4}$$

де $\alpha_\beta = gh(\beta)^{2/(\beta-1)}$.

У якості ілюстрації цього твердження на рис. 1 наведено профілі для 230-мірної випадкової q-арної решітки для $\beta = 2{,}10{,}20{,}30{,}40{,}50$.

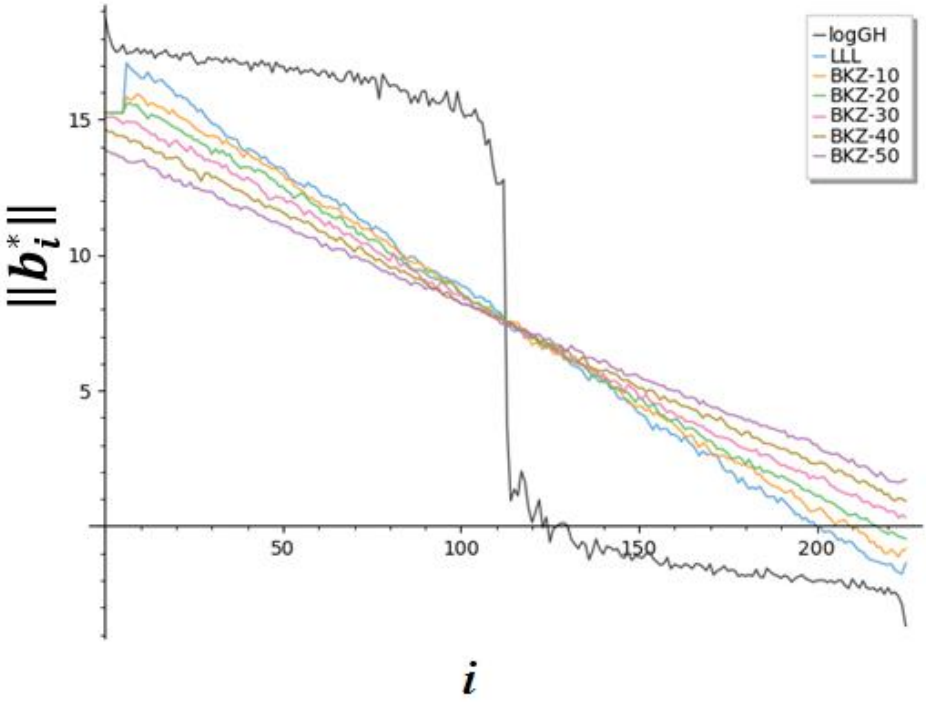


Рис. 1. Профілі 230-мірної q-арної решітки для $\beta = 2{,}10{,}20{,}30{,}40{,}50$.

Варто зауважити, що у більшості робіт замість α_β використовується інша метрика. Доволі зручною метрикою є так званий кореневий фактор Ерміта [6,7], який для базису B визначається як

$$\delta_\beta = (\|b_0\|/vol(\Lambda)^{1/d})^{1/d} \tag{5}$$

З міркувань вище випливає $\delta_\beta = \sqrt{\alpha_\beta}^{1-1/d}$. У роботі [7] запропоновано асимптотичну оцінку

$$\lim_{\beta\to\infty} \delta_\beta = \left(\frac{\beta}{2\pi e} \cdot (\pi\beta)^{\frac{1}{\beta}}\right)^{\frac{1}{2(\beta-1)}} \tag{6}$$

Модель GSA є корисним, проте доволі грубим наближенням форми базису редукованої решітки. По-перше, GSA не враховує того факту, що останній блок буде фактично HKZ редукованим. Оцінити форму HKZ базису можливо аналогічно:

$$h_i = lgh(d-i) - \frac{1}{d-i}\sum_{j<i} h_j \tag{7}$$

Узагальнюючи для довільного базису:

$$\begin{aligned} l_i &= \frac{d-1-2i}{2} \cdot log\alpha_\beta + s, \text{ якщо } 0 \le i \le d-\beta \\ l_i &= h_{i-(d-\beta)} + l_{d-\beta} - h_0, \text{ якщо } d-\beta \le i \le d, \end{aligned} \tag{8}$$

де s є нормуючим фактором, таким, що виконується $\sum l_i = logV$.

Окремо варто розглянути ефекти, що виникають під час редукції q-арних решіток. Такі решітки містять вектор (q, 0, …, 0) та усі його перестановки. Типовою формою базису таких решіток є так звана «Z-форма»:

Приклад Z-форми наведено на рис. 2.

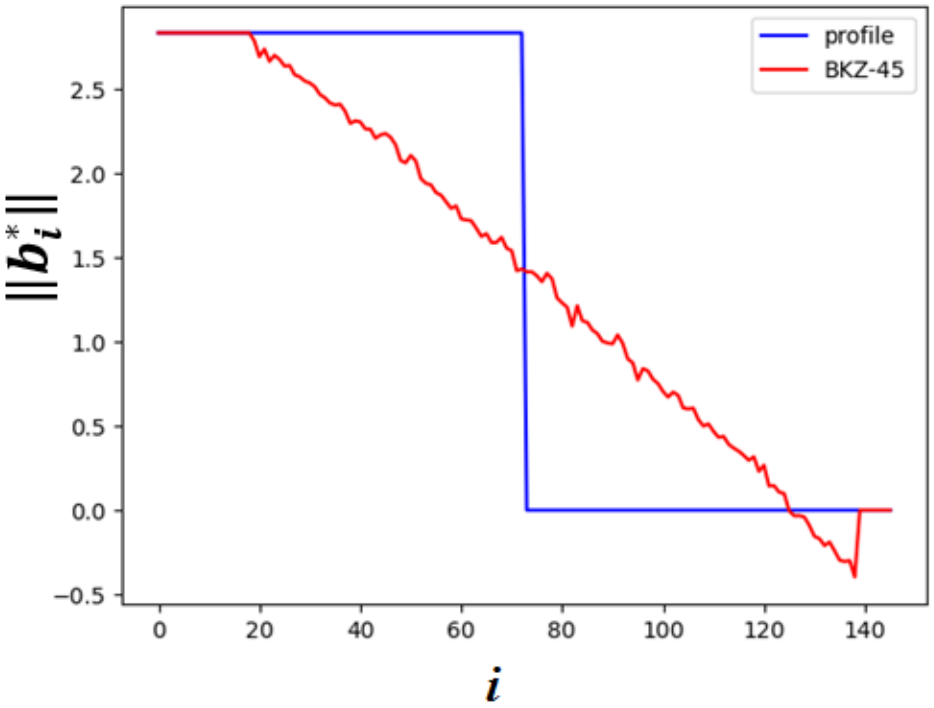


Рис. 2. Z-форма базису решітки

Одним з підходів для моделювання такого базису є модифікація евристики GSA. Такий підхід, зокрема, був популяризований авторами ЕП Crystals-Dilithium. Модифікована евристика GSA (модель ZGSA [6]) визначена наступним чином:

$$\|b_i^*\| = \begin{cases} q, \text{якщо } i \le n-m \\ \sqrt{q} \cdot \alpha_\beta^{\frac{(2d-1-2i)}{2}}, \text{якщо } n-m < i < n+m-1, \\ 1, \text{ якщо } i \ge n+m-1 \end{cases} \quad (9)$$

де $\alpha_\beta = gh(\beta)^{2/(\beta-1)}$ і $m = \frac{1}{2} + \frac{lnq}{2ln\alpha_\beta}$.

Інший підхід базується на основі симуляції. Ідея симуляції ГШ-профілю була запропонована у роботі [7] (Симулятор Чена–Нгуєна). Симулятор замість запуску SVP-оракула визначає очікувану довжину нового вектора за допомогою евристики Гауса. Особливістю симулятора Чена–Нгуєна було використання предобчислених даних для моделювання останнього блоку. У роботі [8] був зала запропонована нова рандомізована версія симулятора Чена–Нгуєна, у якій враховувалася ймовірнісна природа евристики Гауса. Замість точного значення евристики Гауса використовувалися випадкові зміні. Це дало більшу точність симуляції для перших векторів в ГШ-профілі. Втім, для q-арних решіток моделювати «Z-форму» не вдавалося. В роботі [9] запропонований варіант симулятора (симулятор Альбрехта–Лі), що враховує «Z-форму» q-арних решіток.

Для того щоб порівняти якість роботи симуляторів, було проведено ряд експериментів. Для решіток 120, 146, 170 і $q = 17{,}257$ було проведено симуляцію для розмірів блоків 45,50, 55, 60. Для кожного випадку за адаптованою для базисів було обчислено середньоквадратичну похибку. Усереднені значення помилок наведено на рис. 3, 4.

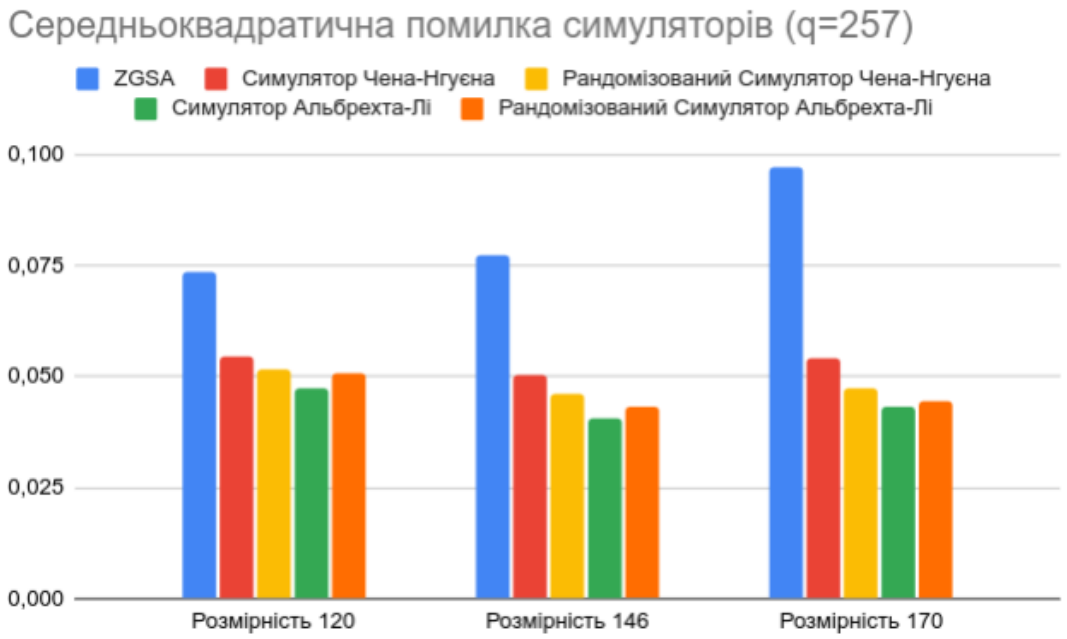


Рис. 3. Середньоквадратичні помилки симуляторів для $q = 256$

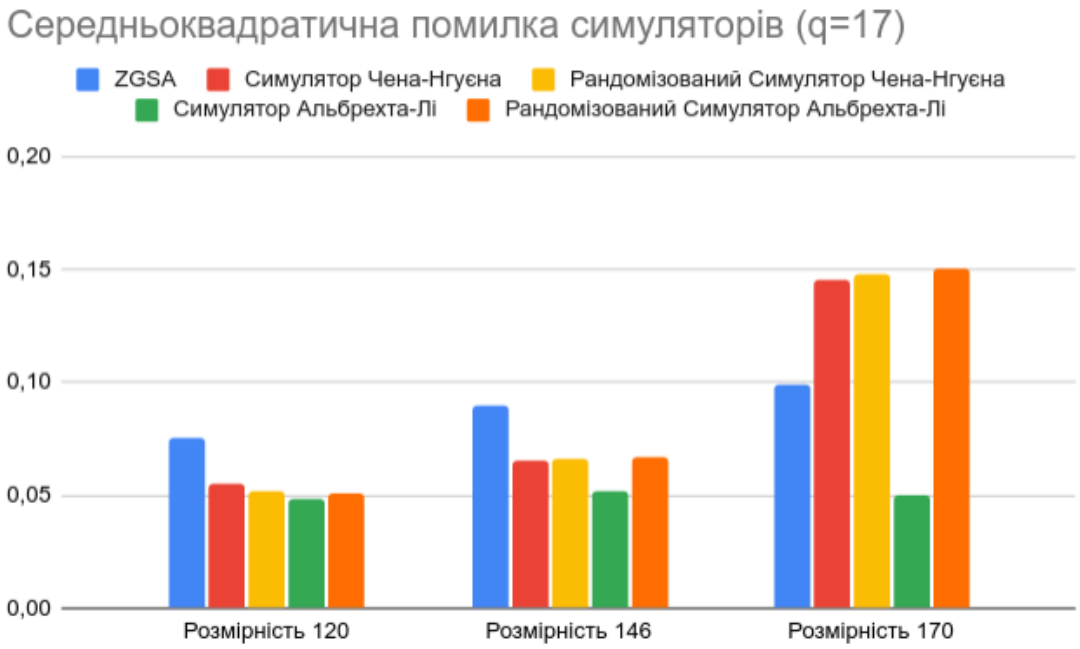


Рис. 4. Середньоквадратичні помилки симуляторів для $q = 17$

З рис. 3, 4 видно, що модель ZGSA доволі непогано себе показує для q=17, проте для q=257 значно програє симуляторам. Симулятор Альбрехта–Лі в обох випадках показує найменшу середньоквадратичну помилку, тож має сенс використовувати його для моделювання редукції решіток. Цікаво, що рандомізована версія симулятора Альбрехта–Лі показує відносно погані результати для q=17 для розмірності 170. Це пов'язано з тим, що через рандомізацію для цієї розмірності якість передбачення Z-форми значно погіршується. На рис. 5 вказано конкретний приклад такої ситуації.

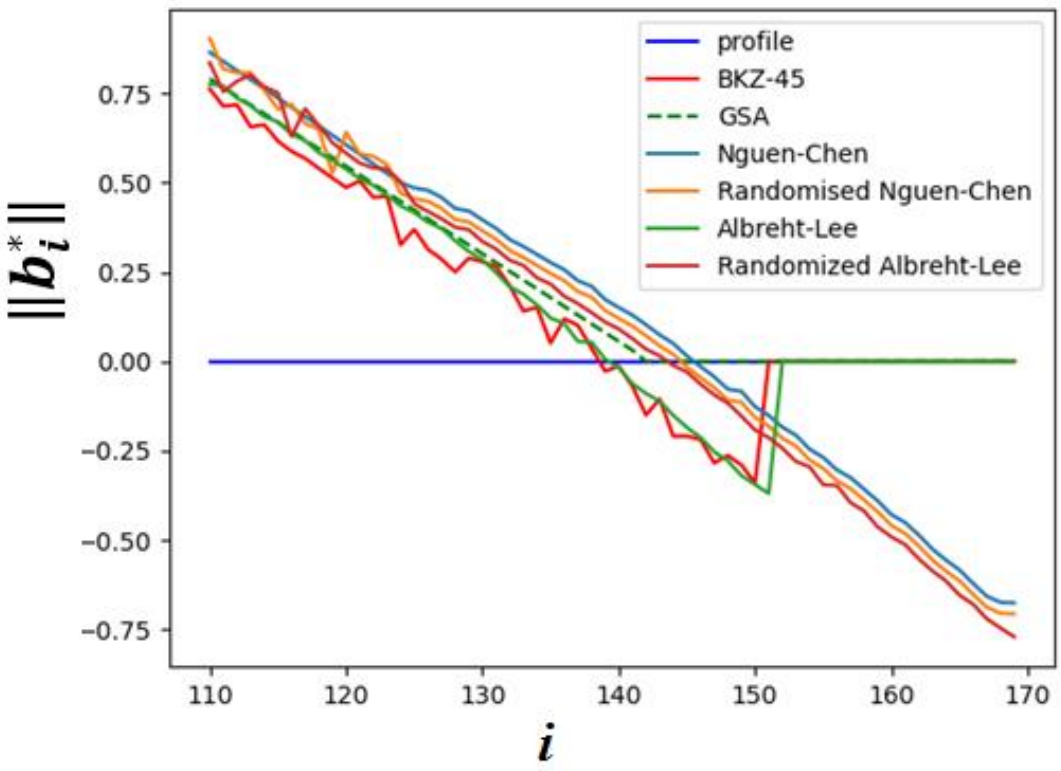


Рис. 5. Приклад поганого передбачення для рандомізованого симулятора Альбрехта–Лі

Тож, найкращу якість передбачення має детермінований симулятор Альбрехта–Лі. Надалі він буде використовуватися для моделювання редукції базису решітки.

3. Вплив розріджених решіток

NTRU решітки окрім q-арної структури мають додаткові алгебраїчні особливості. Особливістю NTRU решіток є велика кількість малих векторів. Зафіксуємо деяке поле $\mathbb{Z}_q[X]/(\phi(X))$ для деякого незвідного полінома $\phi, \deg(\phi) = n$. Якщо $h = g \cdot f^{-1}(mod q)(mod \phi(X))$ для деяких f, g, то решітка розмірності $2n$ з базисом

$$B_{NTRU} = \begin{pmatrix} qI_n & rot(h) \\ 0 & I_n \end{pmatrix} \tag{10}$$

міститиме вектори $(g, f), (x \cdot g, x \cdot f), \dots, (x^{n-1} \cdot g, x^{n-1} \cdot f)$. Ці вектори формують підрешітку розмірності n з базисом

$$B_{NTRU}^{dense} = \begin{pmatrix} rot(g) \\ rot(f) \end{pmatrix} \tag{11}$$

У криптографічному випадку поліноми f, g є малими, тому підрешітка з базисом (11) міститиме велику кількість векторів, що значно менші за евристику Гауса. Знаходження векторів на підрешітці (11) під час редукції решітки дуже швидко призводить до знаходження інших векторів підрешітки, що розбиває редукцію базису (10) на дві незалежні частини і для багатьох параметрів призводить до швидкого знаходження f, g. Приклад такої ситуації наведено на рис. 6.

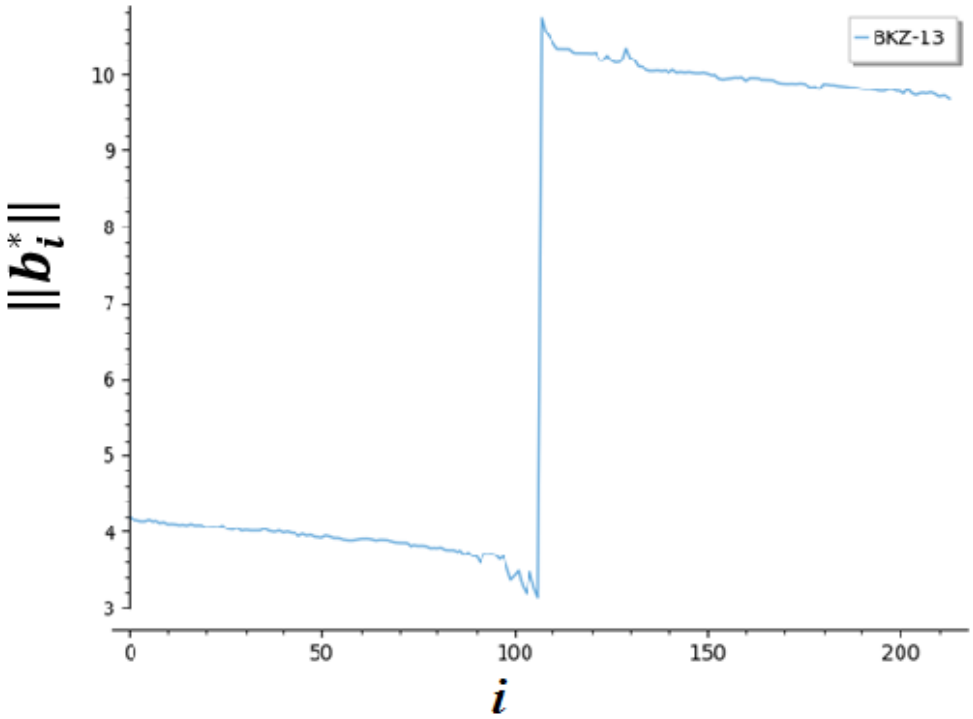


Рис. 6. Перехід на розріджену решітку. Кожна з підрешіток під час редукції не залежить від іншої, що зменшує складність редукції

Вперше ця особливість NTRU решіток була помічена у контексті алгебраїчних атак. Зокрема, так звані атаки на підполе [14]. Проте, у роботі [15] було показано, що розріджені підрешітки призводять до пришвидшення редукції решіток без алгебраїчних технік, що використовувалися в попередніх атаках.

Фактично, при криптоаналізі NTRU решіток є цікавими дві події:

Відновлення таємного ключа – вектор (g, f) буде міститися у базисі решітки.

Потрапляння на розріджену підрешітку – деякий вектор (значно довший за (g, f)) з розрідженої підрешітки буде міститися у базисі

У роботі [15] був отриманий наступний результат. Нехай Λ_{NTRU} – NTRU-решітка розмірності $2n$ з розрідженою решіткою $\Lambda_{NTRU}^{dense} \subset \Lambda_{NTRU}$. Якщо Λ_{NTRU} має базис, що має Z-форму, то під час BKZ-β редукції буде знайдено вектор з розрідженої решітки (що значно більший за таємний ключ), якщо

$$\pi_{n+k-\beta}(v) < \left\| b_{n+k-\beta}^* \right\| \tag{12}$$

де v – найменший вектор на решітці $\Lambda_{proj}^{dense} = \Lambda(\pi_0(b_0), \dots, \pi_0(b_{n+k-1})) \cap \Lambda_{NTRU}^{dense}$.

4. Класифікація та аналіз відомих атак

Відомі атаки на проблему LWE можливо поділити на наступні класи:

- Комбінаторні атаки [16, 17];
- Алгебраїчні атаки [14, 18, 19];
- Атаки декодування [20];
- Атаки розпізнавання [21];
- Атаки вкладення [5];
- Гібридні атаки [22].

До комбінаторних атак належать атака повного перебору та атака MITM (англ. Meet In The Middle) [16]. Зазвичай MITM використовується не самостійно, а як складова більш комплексних гібридних атак. Також до комбінаторних атак можливо віднести атаку BKW [17].

До алгебраїчних атак на LWE відносять атаку Arora-Ge [18], сутність якої полягає у зведенні проблеми LWE до вирішення системи нелінійних рівнянь. Також, якщо розглядати проблему LWE на ідеальних решітках, то існує ряд квантових атак [19], що використовують структуру ідеалів для значного пришвидшення. Проте, такі атаки як правило працюють лише для не криптографічних випадків. Для NTRU алгебраїчною атакою є атака на підрешітку, що була розглянута у розд. 3.

Сутність атак декодування полягає у зведенні проблеми LWE або NTRU до проблеми CVP. Атаки такого роду вимагають побудови та редукції базису решітки таким чином, щоб було можливо вирішити проблему CVP для шуканого таємного вектора.

Атаки вкладення є одними з найбільш ефективних атак на LWE та NTRU. Сутність таких атак полягає у побудові решіток спеціального вигляду, найменший вектор яких містить шуканий секрет. Такі атаки ще називають атаками вкладення, або первинними (англ. Primal) атаками.

Атаки розпізнавання часто називають дуальними атаками через те, що вони зводяться до редукції дуальних (відносно атак вкладення) решіток. У таких атаках використовуються статистичні методи аналізу. Перед криптоаналітиком стоїть задача відрізнення двох розподілів. У поєднанні з комбінаторними методами дуальні атаки можуть давати гарні результати для проблеми LWE. Також їх можливо використовувати для вирішення проблеми SIS.

Гібридні атаки поєднують комбінаторні методи криптоаналізу з атаками вкладення або атаками розпізнавання (гібридні дуальні атаки). Такі атаки при використанні розріджених секретів часто є найкращими для багатьох криптографічних систем.

Розглянемо детально кожен клас атак для оцінки впливу моделей редукції решіток на безпеку криптографічних перетворень.

5. Атаки вкладення

Атака ґрунтується на тому факті, що решітка

$$\Lambda_\omega = \left(x \in \mathbb{Z}^{m+n+1} : \left(I_m \left|\frac{1}{\omega}A\right| - \frac{1}{\omega}b\right)x = 0 mod q\right) \tag{13}$$

містить найменший вектор $v = \lambda_1(\Lambda_\omega) = (e|\omega \cdot s|\omega)$, де ω – параметр масштабування. Значення цього вектора задовільнять LWE рівнянню. Параметр масштабування може бути корисним у випадку, якщо розподіли e та s відрізняються. Типовим значенням є $\omega = \sigma_e/\sigma_s$, де σ_e, σ_s є середньоквадратичними відхиленнями розподілів ймовірностей векторів e, s.

Для оцінки складності пошуку вектора $\lambda_1(\Lambda_\omega)$ не можна застосовувати стандартні припущення на основі Евристики Гауса, оскільки вектор $\lambda_1(\Lambda_\omega)$ набагато менший за $GH(\dim(\Lambda_\omega))$ для типових криптографічних параметрів. На практиці себе добре зарекомендував [5, 6] критерій

$$\left\|\pi_{d-\beta+1}(v)\right\| \le \left\|b^*_{d-\beta+1}\right\|. \tag{14}$$

Ідея, що лежить за критерієм (14), полягає у наступному. Малі вектори у алгоритмі BKZ знаходяться за допомогою SVP-оракула і далі за допомогою алгоритму LLL (більш конкретно – за допомогою кроку редукції за розміром) ці малі вектори вставляються в новий базис. Якщо задано найменше β, для якого виконується (14), то, скоріш за все, потрібний вектор буде знайдено під час останнього виклику SVP-оракула. Знайдений вектор буде вставлений у базис тільки у тому випадку, якщо (14) виконується, за визначенням алгоритму BKZ.

У роботі [23] було доведено, що ймовірність відновлення таємного вектора, якщо виконується вимога (14), складає

$$p = \sum_{i=1}^{d-\beta} \Pr\left[\|\pi_i(v)\| < \min\{\|\pi_i(v) + b_i^*\|, \|\pi_i(v) - b_i^*\|\}\right]. \tag{15}$$

При $\beta > 50$ ймовірність (15) швидко наближається до 1, тож для криптографічних параметрів можливо вважати, що ймовірність відновлення таємного вектора є 1, якщо виконується умова (14). На малих розмірностях іноді таємний вектор може відновлюватися за менших значень β. Це явище пояснюється геометрією решіток [23] і на великих розмірностях не спостерігається, тому при подальшому аналізі ігнорується.

Якщо припустити, що таємний вектор є однорідним (для криптографічних параметрів це є природнім припущенням), то $\left\|\pi_{d-\beta+1}(v)\right\| \approx \sqrt{\beta/d}\|v\|$. У свою чергу, незалежно від розподілу вектору v, з центральної граничної теореми маємо: $\|v\| \approx \sigma\sqrt{d}$. Тож, $\left\|\pi_{d-\beta+1}(v)\right\| \approx \sqrt{\beta}\sigma$, де σ – математичне очікування для компонентів вектора v. Цей факт може бути використаним для перевірки формули (14) на реальних параметрах.

Обчислення правої частини нерівності залежить від моделі редукції решіток. Якщо використовується модель GSA, то

$$\left\|b^*_{d-\beta+1}\right\| \approx \delta_\beta^{2\beta-d} \cdot vol(\Lambda_\omega)^{\frac{1}{d}} = \delta_\beta^{2\beta-d} \cdot q^{\frac{m}{d}}\omega^{(n+1)/d} \tag{16}$$

У випадку використання симуляторів $\left\|b^*_{d-\beta+1}\right\|$ має бути обчислено експериментально.

Варто зауважити, що складність атаки залежить не монотонно від значення m. У роботі [21] доведено, що оптимальним значенням є

$$m_{opt} = \left\lceil \sqrt{\frac{(n+1)(logq-log\omega)}{log\delta_\beta}} - (n+1) \right\rceil \tag{17}$$

Таким чином, оцінка атаки вкладення зводяться до знаходження найменшого β, для якого виконується (14). При цьому мають бути задані:

- Параметри решітки (n, q, σ)
- Модель редукції решіток, що визначає $\left\|b^*_{d-\beta+1}\right\|$
- Модель часу роботи редукції

Модель часу роботи редукції решіток слугує для перетворення оптимальних параметрів редукції у конкретну оцінку безпеки, тому при дослідженні впливу моделей редукції решіток на безпеку, її можливо не задавати, а досліджувати безпосередньо оптимальні параметри атаки.

На рис. 7 наведено результати моделювання атаки вкладення з використанням моделі GSA та симулятора Альбрехта–Лі для $n = 256$ при $\sigma = \{1.1, 1.3, 1.5, 1.8, 2.1\}$. З рис. 7 видно, що при збільшенні параметра q зменшується оптимальне для атаки значення β, що є логічним, враховуючи формулу (16): при збільшенні q права частина рівняння (14) збільшуватиметься, у той час як ліва частина залишатиметься такою ж. Втім, графік для симуляторів має дві ключові відмінності. Відрізняється сила впливу параметра σ на оптимальне значення β. При використанні моделі GSA спостерігається трохи більший розмах між випадками $\sigma = 1.1$ та $\sigma = 2.1$ на більшій частині досліджуваного простору параметрів. При малих значеннях q це може суттєво впливати на значення β.

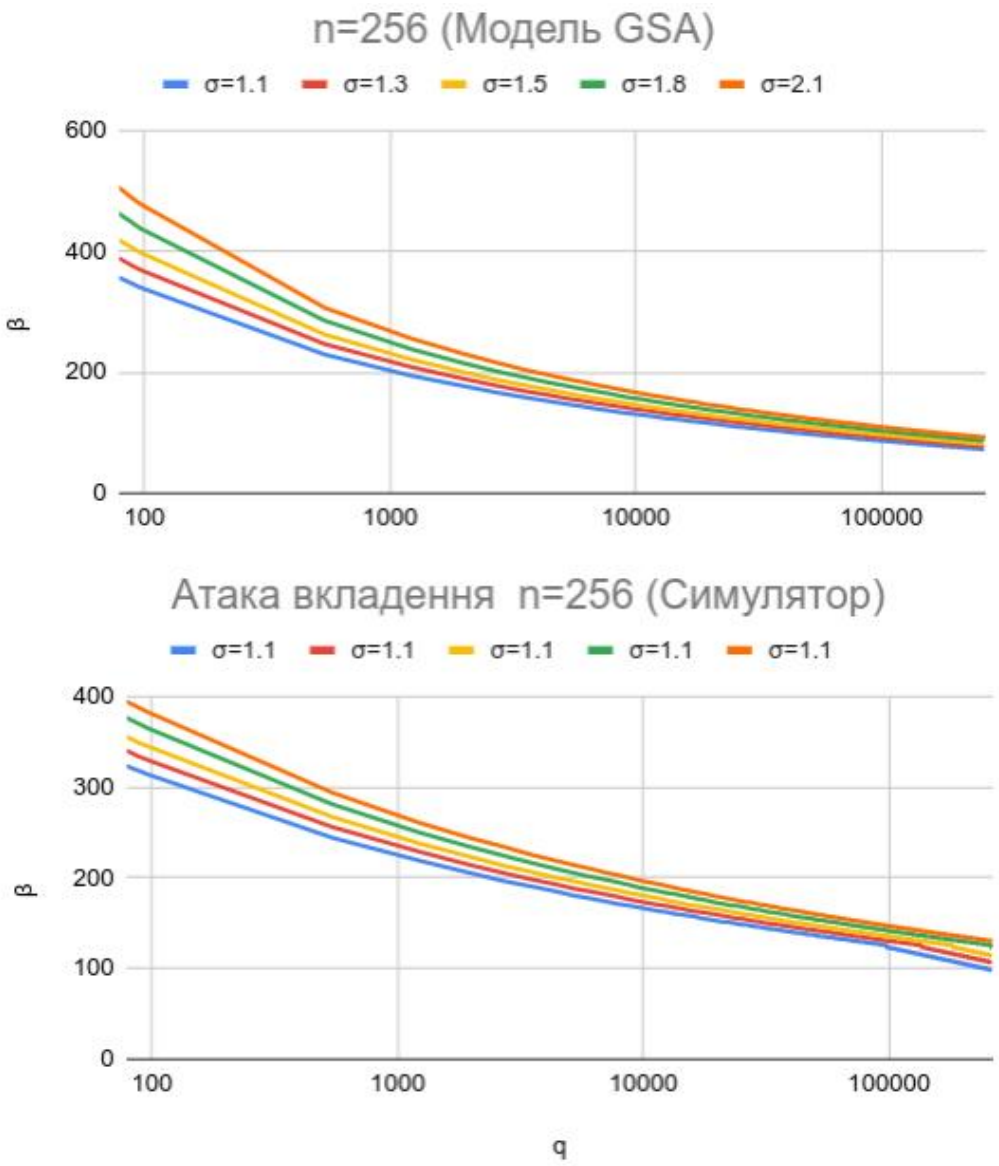


Рис. 7. Результати моделювання атаки вкладення

Для більш наглядної демонстрації різниці між симулятором та моделлю GSA на рис. 8 наведено накладені один на одного графіки з рис. 7. З рис. 8 видно, що на малих значеннях q симулятор дає менші значення параметра β, у порівнянні з моделлю GSA. У той же час, для великих значень q симулятор дає більші значення β, що свідчить про те, що модель GSA дещо занижує рівень безпеки. Отримані свідчення можливо пояснити тим, що останній блок у профілі решітки буде HKZ редукованим, а отже $\left\|b^*_{d-\beta+1}\right\|$ буде мати трохи більше значення, ніж передбачене GSA значення. Оскільки симулятор враховує це явище, то данні на рис. 8 виглядають цілком природніми.

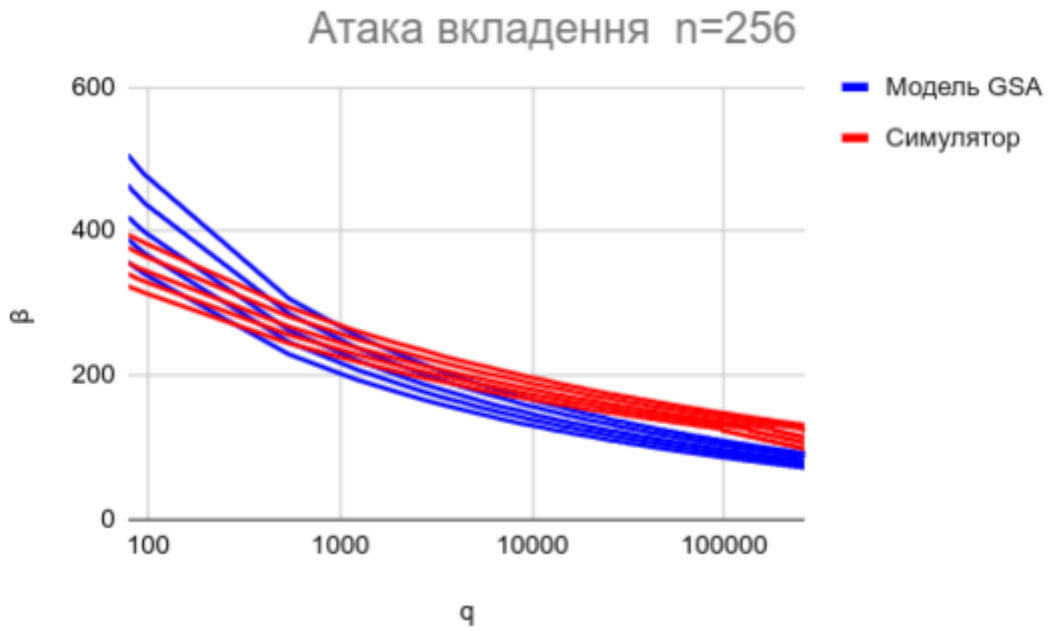


Рис. 8. Порівняння моделі GSA та симулятора Альбрехта–Лі

При збільшенні параметра n результати моделювання виглядають схожим чином і для них зберігаються усі описані явища. На рис. 9 наведено результати моделювання для $n = 256,512,1024,2048$ при фіксованому $\sigma = 1.1$.

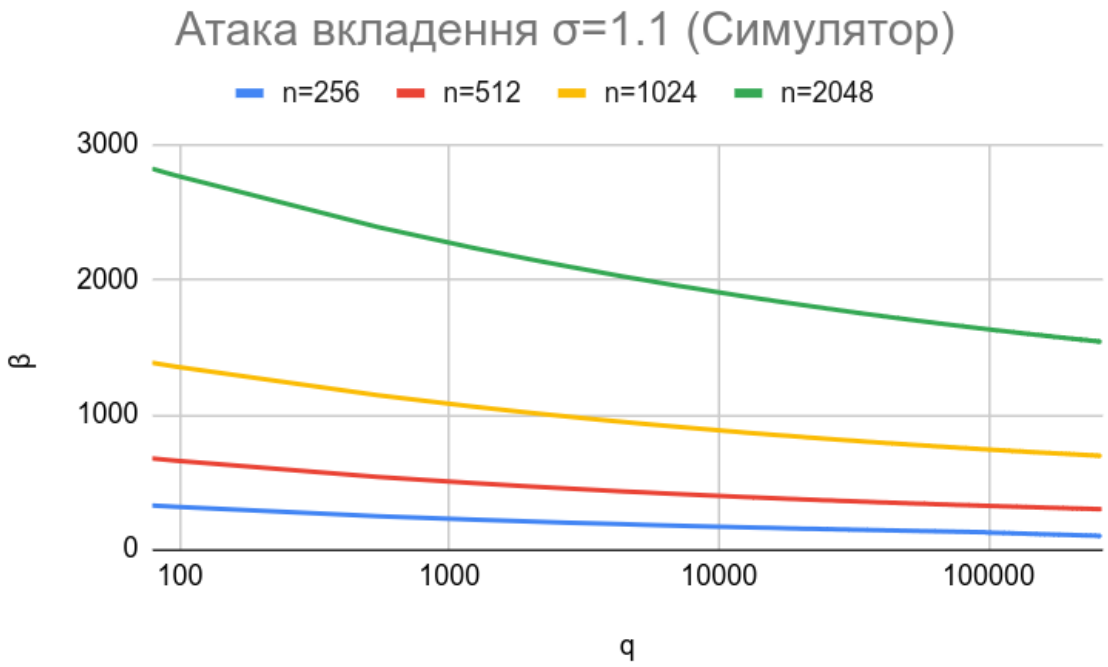


Рис. 9. Результати моделювання атаки вкладення
для $n = 256,512,1024,2048$ при фіксованому $\sigma = 1.1$

З рис. 9 видно, що збільшення n дає лінійний приріст значення β, з чого випливає стратегія пошуку оптимальних загальносистемних параметрів: зафіксувати значення n, що дає близьке до необхідного рівня безпеки значення і підлаштовувати рівень безпеки змінюючи параметри q, σ.

Якщо решітка має розріджену підрешітку, то необхідно додатково використовувати формулу (12) для урахування можливості переходу на розріджену підрешітку. Оскільки формула (12) вже враховує Z-форму базису, то її не має сенсу застосовувати в моделі GSA. Для NTRU решіток починаючи з $q \approx 0.0038 \cdot n^{2.484}$ відбуватиметься перехід на розріджену решітку. При обчисленні безпеки параметрів це необхідно враховувати.

6. Атаки декодування

Атака декодування зводить проблему LWE до задачі знаходження найближчого вектора. Для цього будується решітка

$$\Lambda = \{x \in \mathbb{Z}^m | x = A \cdot s \bmod q\} \tag{18}$$

і для вектора $t = A \cdot s + e$ вирішується задача пошуку найближчого вектора $v = A \cdot s$, знаючи який можливо легко відновити (s, e). Щоб знайти найближчий вектор необхідно провести редукцію базису (18). У загальному випадку складність атаки можливо знайти за формулою

$$(T_{red} + T_{cvp})/p_{succ} \tag{19}$$

де T_{red} – час редукції базису решітки (18), T_{cvp} – час алгоритму пошуку найближчого вектора, p_{succ} – ймовірність вдалого завершення атаки.

У якості алгоритму пошуку найближчого вектора, як правило, використовується алгоритм Бабаї та його узагальнення. Алгоритм Бабаї гарантує [22], що $v - t \in P_{\frac{1}{2}}(B^*)$, тож для цього випадку необхідною умовою вдалого завершення атаки є $e \in P_{\frac{1}{2}}(B^*)$.

У випадку нормального розподілу помилки, у роботі [20] запропонована наступна оцінка:

$$p_{succ} = \Pr\left[e \in P_{\frac{1}{2}}(B^*)\right] = \prod_{i=1}^{m} erf\left(\frac{\|b_i^*\|\sqrt{\pi}}{2s}\right) \tag{20}$$

Для класичного алгоритму Бабаї T_{cvp} є поліноміальним і може не враховуватися. Проте, використання більш складних алгоритмів пошуку найближчого вектора, що працюють за субекспоненційний час, може зменшити (19). Такі алгоритми, як правило, тісно повязані з SVP оракулами. Ідеї, що використовуються для побудови SVP оракулів, також можуть бути адаптованими для побудови алгоритмів вирішення задачі CVP, як як це було показано у роботі [20]. Сутність ідеї полягає у тому, щоб у алгоритмі Бабаї обчислювати не тільки найоптимальніші координати $c_0, \dots, c_{d-1}$, а для кожної координати c_i перебирати d_i найоптимальніших значень. Тоді, складність алгоритму пошуку найближчого вектора при використанні найпростішої стратегії перебору складатиме $O(\prod_i d_i)$, яка вже буде не поліноміальною. Тоді, оцінка (19) буде мінімізуватися, коли $T_{red} \approx T_{cvp}$. Ймовірність p_{succ} відповідно складатиме

$$p_{succ} = \prod_{i=1}^{m} p_{succ}^i = \prod_{i=1}^{m} erf\left(\frac{d_i\|b_i^*\|\sqrt{\pi}}{2s}\right) \tag{21}$$

Питання вибору значень d_i в літературі є малодослідженим. Якщо використовувати модель GSA, то значення $\|b_i^*\|$ будуть розподілені за експоненціальним розподілом. Значення d_i, що мають відмінне від одиниці значення, будуть згруповані у хвості профіля решітки. Проте, оскільки у q-арних решіток останні значення $\|b_i^*\|$ сильно відхиляються від GSA і $\|b_i^*\| \approx 1$, що повинно зменшувати відповідні значення d_i для цих векторів, а отже і зменшувати відповідне значення T_{cvp}.

Для пошуку відповідних значень d_i зафіксуємо мінімальні ймовірність успішного виконання атаки p_0. Стратегія пошуку значень d_i випливає з міркувань, що існує така ймовірність p_{avg}, що

$$p_0 \le \prod_{i=1}^{m} p_{succ}^{i} < \left(p_{avg}\right)^{m} \tag{22}$$

Зрозуміло, що мінімальне таке значення є $p_{avg} = (p_0)^{1/m}$. Тоді, якщо кожне p_{succ}^{i} буде більшим за p_{avg}, то буде виконуватися $p_0 \le p_{succ}$. Звідси, маємо

$$p_{succ}^{i} = erf\left(\frac{d_i \|b_i^*\|\sqrt{\pi}}{2s}\right) \le p_{avg} \Rightarrow$$
$$d_i \le \frac{erfinv(p_{avg})\cdot 2s}{\|b_i^*\|\sqrt{\pi}} \tag{23}$$

На рис. 10 зображена залежність p_{avg} від p_0 для різних значень параметра m.

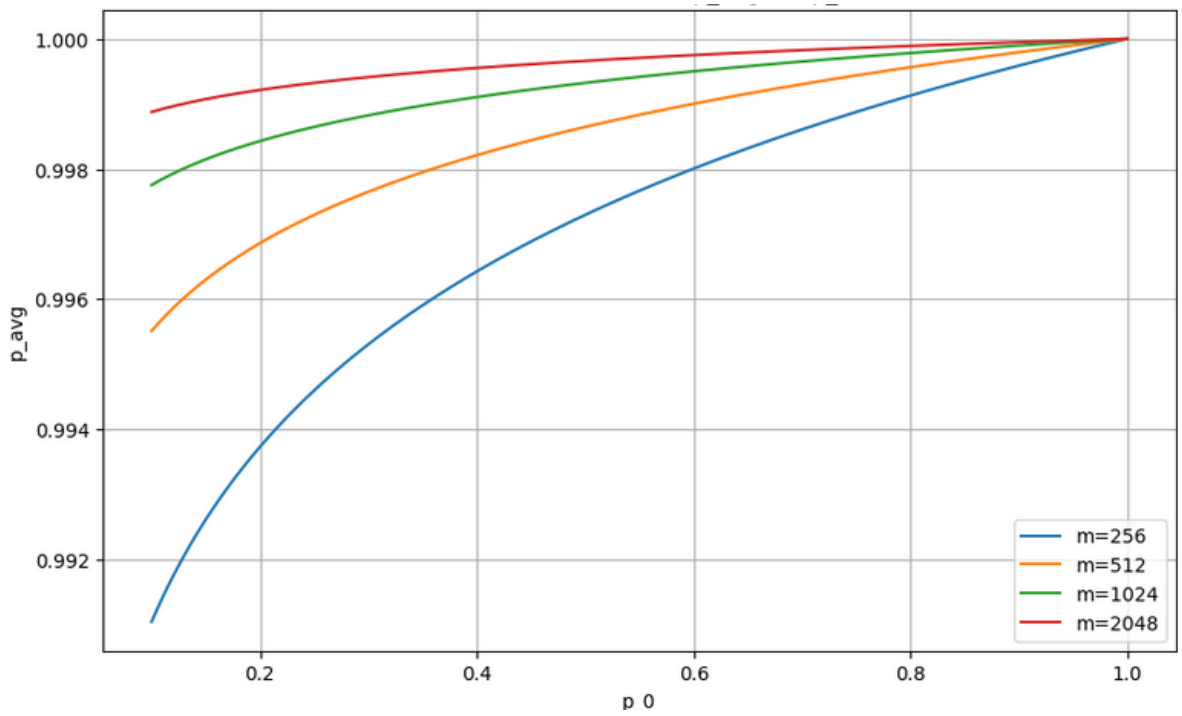


Рис. 10. залежність p_{avg} від p_0 для різних значень параметра m.

З рис. 10 видно, що для типових значень параметра m значення параметра p_{avg}, лежить близько до 1, тобто вплив ймовірності p_0 на значення d_i є не значним.

Тож, для атаки декодування мають бути задані

- Модель редукції решіток, що визначає профіль решітки
- Модель часу роботи редукції
- Модель часу роботи алгоритму CVP

У межах дослідження було проведено моделювання атаки декодування для моделі GSA та симулятора Альбрехта–Лі для розмірностей $N = 256,512,1024,2048$ та значень параметра $\sigma = 1.1,1.3,1.5$. Результати моделювання для розмірності 512 наведені на рис. 11.

З рис. 11 видно, що вартість атаки з використанням симулятора Альбрехта–Лі є вищою, проте такої великої переваги, як в атаках вкладення не має. Це пояснюється тим, що вартість атак вкладення залежить лише від значення $\left\|b_{d-\beta+1}^{*}\right\|$, у той час, як вартість атак декодування залежить цілком від всього профіля редукованої решітки. Також видно, що складність етапу редукції решітки та етапу пошуку найближчого вектора хоч і наближаться один до од-

ного, проте не дорівнюють один одному, тобто мінімум формули (19) не досягається через дискретність параметрів.

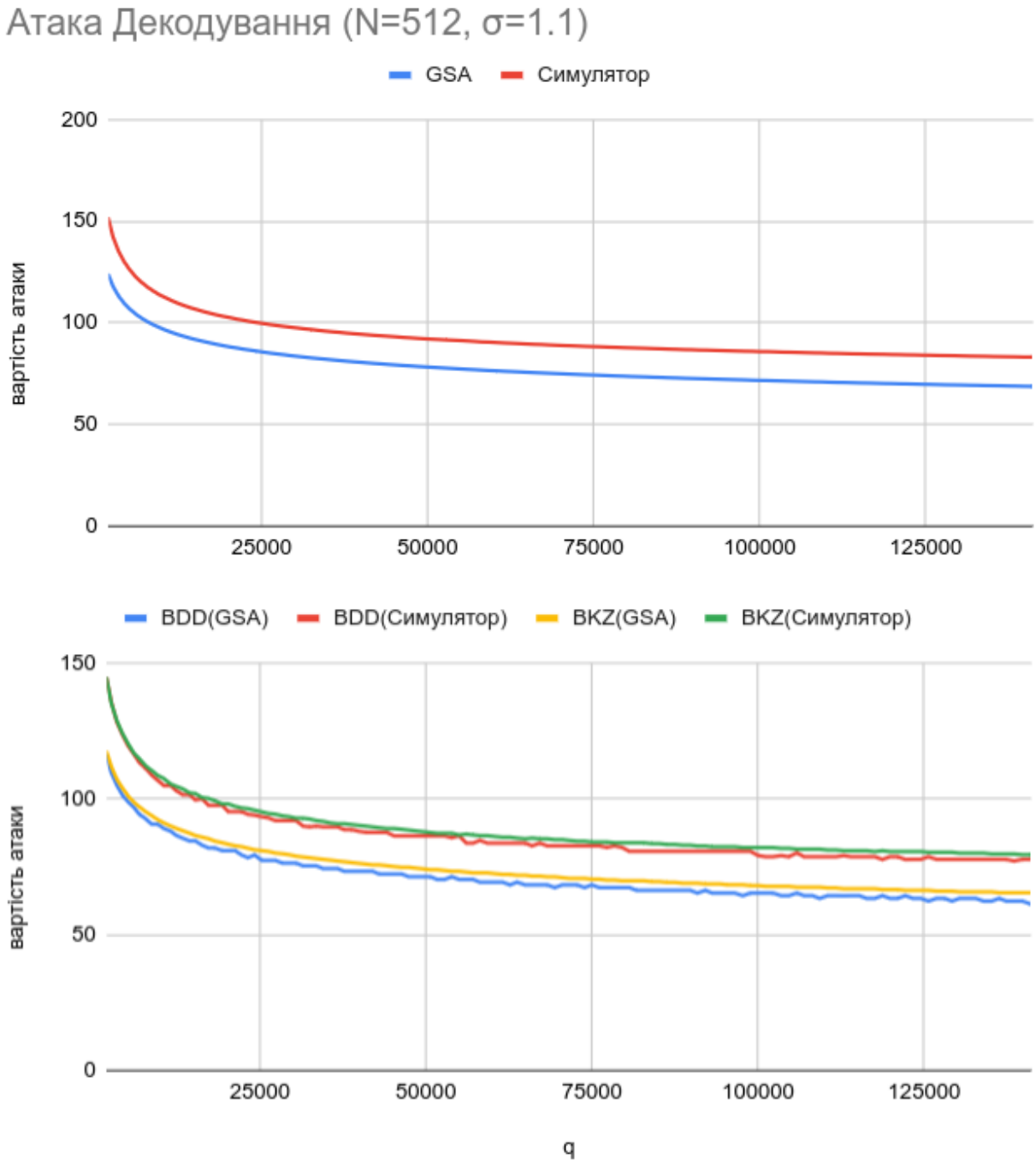


Рис. 11. Результати моделювання атаки декодування для $N = 512, \sigma = 1.1$

На рис. 12 показано вплив збільшення параметра N на складність атак декодування.

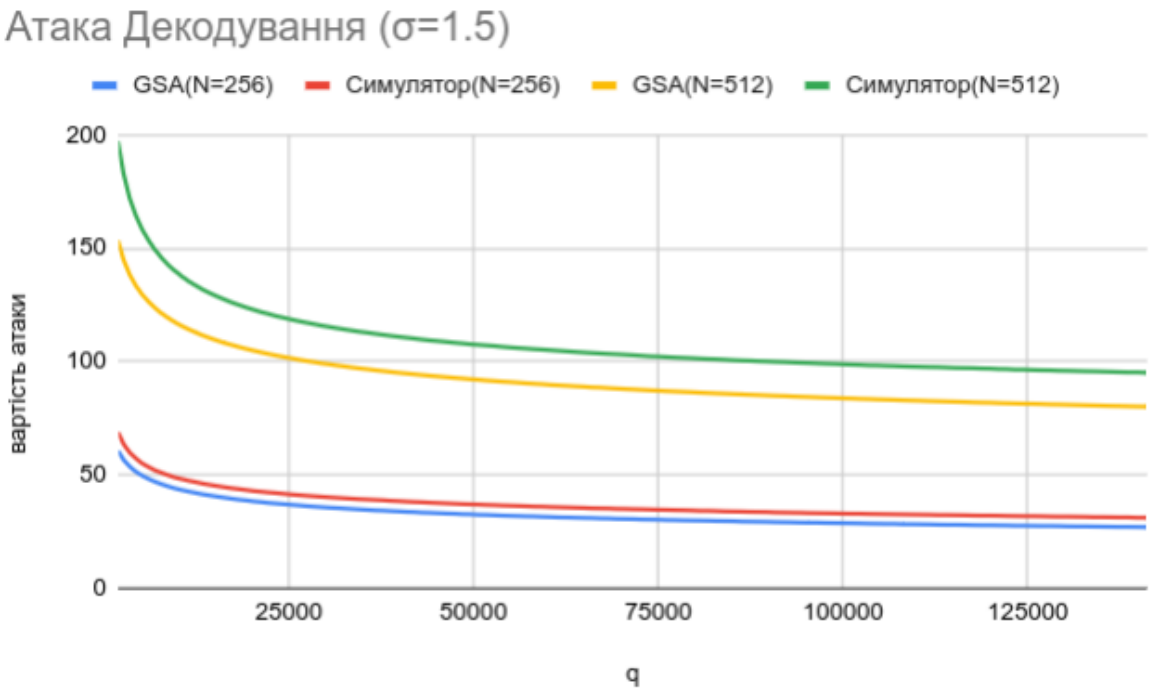


Рис. 12. Вплив параметра N на складність атак декодування

Як видно з рис. 12, збільшення параметра N вдвічі збільшує складність атаки приблизно вдвічі. Цікаво, що при цьому зростає вплив симулятора на складність атаки, чого так явно не спостерігалося для атак вкладення, що , знов ж таки, пояснюється тим, що форма профіля базису в атаках декодування впливає сильніше на складність атаки.

На рис. 13 показано вплив параметра σ на вартість атак декодування.

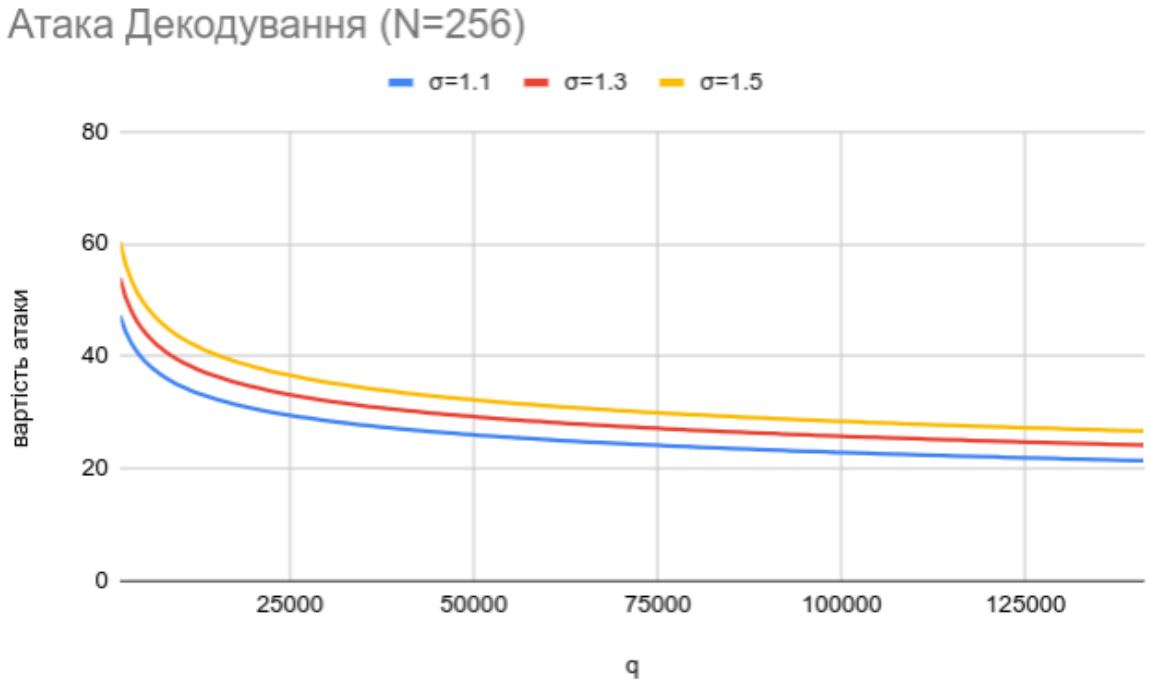


Рис. 13. Вплив параметра σ на вартість атак декодування

Великої різниці у впливі параметра σ при використанні моделі GSA та симулятору не має. Як і в атаках вкладення, параметр σ можливо використовувати для уточнення параметрів безпеки.

На рис. 14 наведено порівняння складності атак вкладення та декодування.

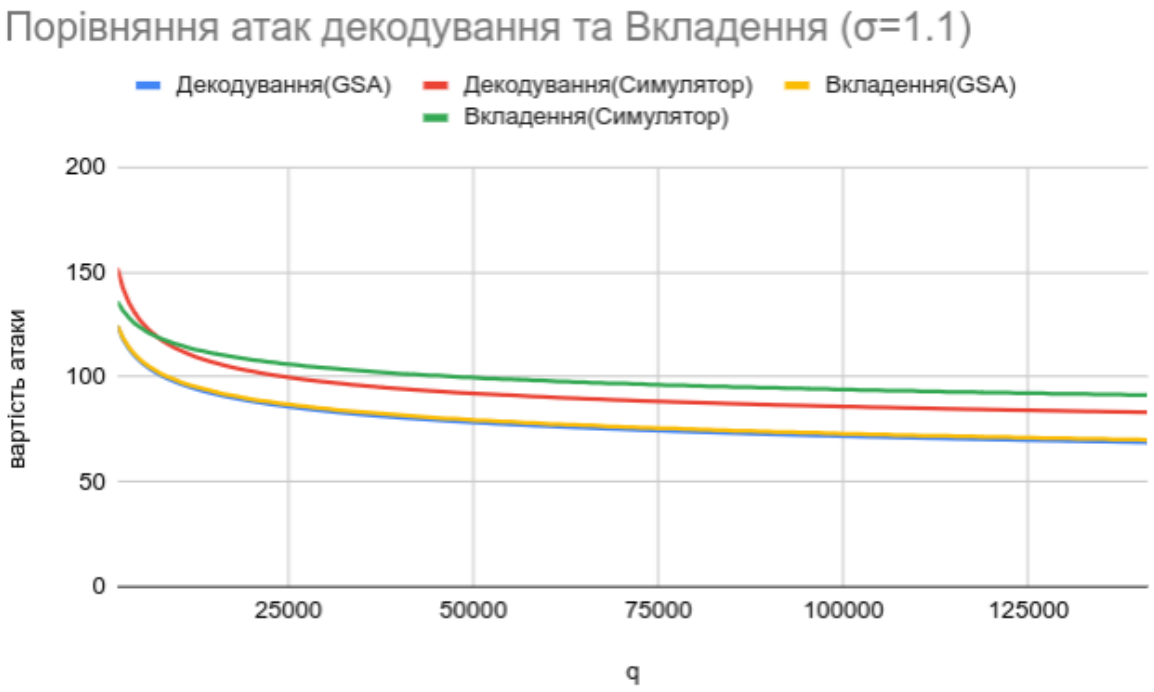


Рис. 14. порівняння атак вкладення та декодування

З рис. 14 можливо зробити декілька висновків. При використанні моделі GSA складність атак декодування та вкладення є майже однаковою. Фактично, різниця настільки не суттєва, що їх вартість можливо вважати однаковою. Проте, при використанні симуляторів різниця між атаками вкладення та декодування стає помітною. Для переважної частини параметрів атака декодування перевершує атаку вкладення, проте на малих значеннях параметра q атака вкладення все ж стає кращою. Різниця є достатньо малою, проте при виборі параметрів, все ж, її варто враховувати. Тож, при оцінці безпеки не можна нехтувати атаками декодування.

7. Атаки розпізнавання

Атаки розпізнавання є статистичними атаками, у яких супротивник намагається відрізнити пари $(A, t = A^T s + e)$ від пари (A, b) з рівномірного розподілу. Це можливо зробити, якщо відомо деякий малий вектор v, для якого виконується $Av = 0 mod q$ (тобто він лежить на дуальній решітці $\Lambda_q^{\perp}(A^T)$). Для вектора $t = A^T s + e$ скалярний добуток $\langle v, t \rangle mod q$ буде мати нормальний розподіл, оскільки $\langle v, t \rangle = vA^T s + \langle v, e \rangle mod q = \langle v, e \rangle$. Для вектору b відповідний розподіл буде рівномірним.

Атаки розпізнавання у багатьох моделях безпеки, наприклад у Crystals-Dilithium, чи New-Hope, аналізуються чисто як атаки, що дозволяють відрізнити розподіл LWE від рівномірного розподілу. Звичайно, факт відрізнення LWE розподілу від рівномірного руйнує усі докази безпеки у таких моделях безпеки, як IND-CCA для схем асиметричного шифрування, чи IND-CMA для електронних підписів, проте в реальному світі зловмисників цікавить саме відновлення таємного ключа, а не лише відрізнення розподілів. Перетворення атаки розрізнення на атаку відновлення ключів потребує деяких додаткових обчислювальних ресурсів, тому такі оцінки є дещо заниженими. Враховуючи, що навіть такі занижені оцінки є гіршими за атаки вкладення та декодування, то уточнені оцінки будуть ще гіршими. Тож, у моделі безпеки їх можливо не враховувати.

8. Гібридні атаки

Основна ідея гібридної атаки ґрунтується на тому, що якщо q – просте число, то для будь-якої вимірної -арної решітки Λ базис можливо представити у вигляді

$$B = \begin{pmatrix} B_1 & B_2 \\ 0 & I_r \end{pmatrix} \in \mathbb{Z}^{n \times n}, \tag{24}$$

де $0 < r < n$, $B_1 \in \mathbb{Z}^{(n-r)\times(n-r)}$, $B_2 \in \mathbb{Z}^{(n-r)\times r}$.

Використовуючи структурованість базиса, довільний вектор $v \in \Lambda$ можливо представити як конкатенацію векторів меншої розмірності:

$$v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = B \begin{pmatrix} x \\ v_2 \end{pmatrix} = \begin{pmatrix} B_1 x + B_2 v \\ v_2 \end{pmatrix} \tag{25}$$

$$v_1 \in \mathbb{Z}^{(n-r)}, v_2 \in \mathbb{Z}^r$$

для деякого $x \in \mathbb{Z}^{(n-r)}$.

З формули (25) маємо рівняння $B_2 v_2 = -B_1 x + v_1$. Оскільки v_1 є малим вектором, то вектор $B_2 v_2$ знаходиться близько до решітки $\Lambda(B_1)$ і за умови, що B_1 є достатньо редукованим базисом, може бути відновлений за допомогою алгоритма найближчої площини Бабаї [22]: $v_1 = NearestPlane_{B_1}(B_2 v_2)$.

Гібридна атака складається з трьох етапів:

- Редукувати базис B_1.
- Знайти вектор v_2 за допомогою комбінаторних технік.
- Знайти вектор v_1 за допомогою алгоритма $NearestPlane$

Оскільки розмірності, у яких виконується пошук, є меншими, то загальний час виконання буде значно меншим за умови, що комбінарторна частина атаки має не велику оцінку. Оскільки реалізація комбінаторної частини атаки сильно залежить від конкретної криптографічної схеми, то конкретні оцінки комбінаторної частини атаки будуть сильно відрізнятися для різних схем. Фактично, гібридна атака є модифікацією атаки декодування, тому графік для простору параметрів буде схожим на рис. 11, тільки зміщеним на деякий фактор, який визначається комбінаторною частиною.

Є цікавим той момент, що формула (21), яка визначає ймовірність знаходження вектора v_1, отримана з припущення, що таємний вектор матиме нормальний розподіл, у той час, як гібридна атака застосовується переважно для векторів, що мають розподіл відмінний від нормального. Ця проблема у літературі зазвичай обходиться стороною.

Ми пропонуємо при оцінці гібридної атаки для розподілів, що відмінні від нормального, апроксимувати цей розподіл ймовірностей нормальним розподілом, що мінімізує відстань Колмогорова-Смірнова, яка визначена як максимальна абсолютна різниця між двома емпіричними функціями розподілу.

Для нормального розподілу у межах $[-\epsilon, +\epsilon]$ для $\epsilon = 1, \ldots, 10$ ми обчислили оптимальні значення дисперсії σ у табл. 1. На рис. 15 у якості ілюстрації наведено порівняння функцій розподілу для $\epsilon = 3$.

Таблиця 1

Оптимальні значення σ для мінімізації відстані Колмогорова-Смірнова

ϵ	Оптимальне σ	Відстань Колмогорова–Смірнова	ϵ	Оптимальне σ	Відстань Колмогорова–Смірнова
1	1.03	0.16666	6	4.14	0.07339
2	1.71	0.12074	7	4.73	0.06963
3	2.30	0.09644	8	5.34	0.06701
4	2.93	0.08589	9	5.94	0.064911
5	3.53	0.07807	10	6.54	0.063163

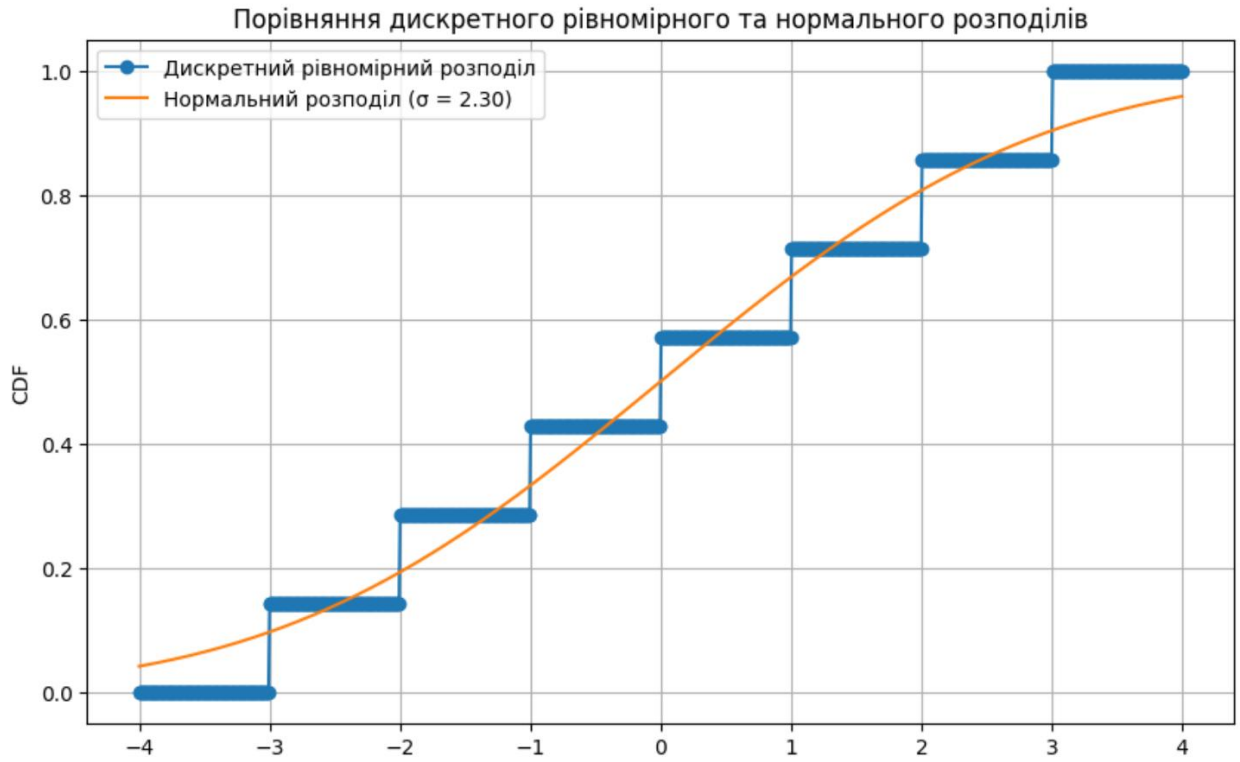


Рис. 15. Порівняння дискретного рівномірного розподілу
та його апроксимації нормальним розподілом

Отримані значення, на нашу думку, дозволяють з достатньою точністю апроксимувати заданий розподіл нормальним розподілом і застосувати описаний вище підхід до атаки декодування та гібридної атаки.

9. Атаки на проблему SIS

Проблема SIS дещо відрізняється від проблем LWE та NTRU тим, що вимагається знаходження вектора, що є малим у l_∞ нормі. Відповідний вектор лежить на решітці

$$\Lambda(A) = \{z \in \mathbb{Z}^d | Az = 0\} \tag{26}$$

Існуючі в літературі підходи до оцінки складності проблеми SIS через редукцію решіток вважають, що нам відомий розмір шуканого вектору у l_2 нормі. Проте, з визначення проблеми не випливає, що нам необхідний вектор саме з конкретною l_2 нормою. Тому ми розробили власний підхід, який враховує те, що для шуканого вектора з l_∞ нормою можуть бути знайдені рішення з різними l_2 нормами.

Щоб оцінити ймовірність події знаходження вектора v, що має l_∞ норму B можливо використати властивість концентрації мери на гіперсфері. Для -вимірної гіперсфери ймовірність того, що довільна компонента вектора v_i буде далеко від середнього значення експоненціально зменшується з збільшенням відстані. Для кожного v_i маємо

$$\Pr[|v_i| \geq B] \leq 2exp\left(-\frac{\left(\frac{B\sqrt{d}}{\|v\|_2}\right)^2}{2}\right) \tag{27}$$

Відповідно, для усього вектора маємо:

$$\Pr\left[\|v\|_\infty \leq B\right] \approx \left(1 - 2exp\left(-\frac{\left(\frac{B\sqrt{d}}{\|v\|_2}\right)^2}{2}\right)\right)^d \tag{28}$$

Запропонований підхід до вирішення SIS полягає у наступному:

- Провести редукцію базису SIS-решітки з параметром β_1
- За допомогою алгоритму просіювання у розмірності β_2 отримати N малих векторів з значенням норми α
- Для заданого значення α знайти ймовірність p_{succ} того, що l_∞ норма не перевищує значення B за формулою (28).
- Оцінити складність атаки як $(T_{red} + T_{sieve})/(\min(1, N \cdot p_{succ}))$

Алгоритм просіювання може повернути н $2^{0.2075\beta}$ векторів з нормою $\alpha = \rho \cdot \|b_0\|_2$, де ρ оцінюється як

$$\sqrt{4/3} \cdot \delta_{\beta_1}^{\beta_1 - 1} \delta_{\beta_2}^{1-\beta_2} \tag{29}$$

Моделювання атаки показало, що застосування симуляторів майже не впливає на оцінки безпеки для атаки на SIS. На рис. 16 показано результати моделювання.

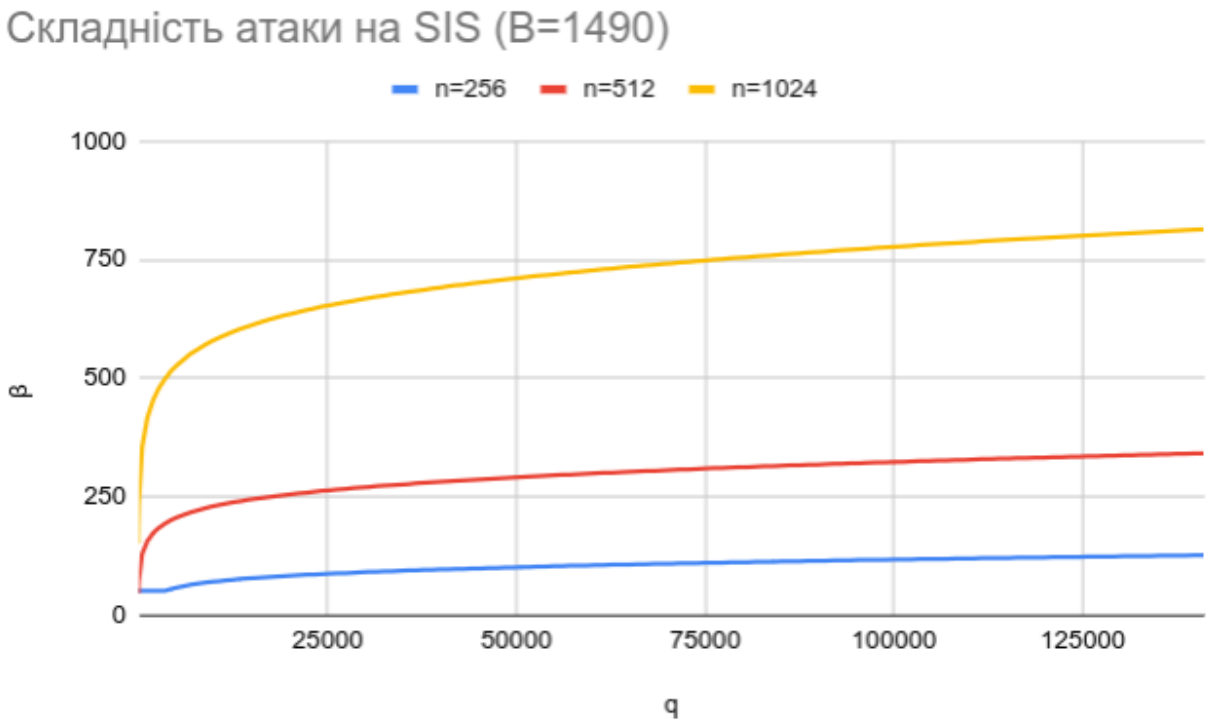


Рис. 16. Результати оцінки атаки на SIS

Додатково для оцінки впливу нової моделі на складність криптоаналізу SIS були розраховані оцінки для Crystals-Dilithium. Відповідні оцінки наведено у табл. 2.

Таблиця 2

Оцінка складності проблеми SIS для Crystals-Dilithium

Рівень безпеки NIST	Оцінка авторів Crystals-Dilithium (біт)	Наша оцінка
2	123	111
3	186	163
5	265	236

З табл. 2 видно, що наша оцінка дає менші значення безпеки, ніж очікувалися авторами Crystals-Dilithium.

Висновки

1. Для виявлення найкращої моделі редукції решіток, що дозволяє найточніше моделювати профіль решітки, було обчислено середньоквадратичну помилку на решітках малої розмірності для моделі GSA (ZGSA) та варіантів симулятору Чена–Нгуєна і симулятору Альбрехта–Лі. Детермінований симулятор Альбрехта–Лі показав найкращі результати для усіх значень параметрів.

2. При врахуванні алгебраїчної структури q-арних решіток в атаках вкладення було виявлено, що модель GSA занижує значення безпеки. Цей ефект пояснюється тим, що GSA не враховує того, що останній блок в базисі є HKZ-редукованим і має іншу форму. На малих значеннях параметра q уточнені оцінки показують менші показники безпеки, проте зі збільшенням праметра ситуація повністю змінюється. Оцінки безпеки стають більшими, ніж для GSA. Більшість існуючих криптографічних параметрів потрапляють у другу зону. Це вказує на те, що існуючи схеми переважно є безпечнішими, ніж вважалося раніше. Для NTRU решіток також необхідно враховувати можливість переходу на розріджену решітку.

3. Для атак декодування було запропоновано стратегію вибору параметрів атаки d_i. З використанням цих параметрів було показано, що атаки декодування можуть бути кращими за атаки вкладення. Вплив алгебраїчної структури q-арних решіток на атаки відновлення є не таким сильним, як при атаках вкладення. Це пояснюється тим, що на атаки декодування впливає вся форма профіля, а не лише конкретне значення в останньому блоці, як у атаках вкладення.

4. Гібридні атаки є, фактично, узагальненням атак декодування, хоча вони історично з'явилися раніше. Оцінки гібридних атак та атак декодування ґрунтуються на тому, що розподіл таємного вектора є нормальним. Проте, це не так для більшості параметрів, для яких гібридні атаки можливо застосувати. Щоб подолати цю ситуацію було запропоновано апроксимувати відповідні розподіли нормальним розподілом, мінімізуючи відстань Колмогорова-Смірнова та обчисленні конкретні оптимальні параметри апроксимуючих нормальних розподілів.

5. Існуючі в літературі підходи до оцінки складності проблеми SIS через редукцію решіток вважають, що нам відомий розмір шуканого вектору у l_2 нормі. Проте, з визначення проблеми не випливає, що нам необхідний вектор саме з конкретною l_2 нормою. У межах дослідження було розроблено метод, що враховує факт того, що при фіксованих вимогах до l_∞ норми, l_2 норма може мати різні значення. Отримані оцінки показують, що складність криптоаналізу Crystals-Dilithium є меншою, ніж вважалося раніше. Для п'ятого рівня безпеки різниця становить 2^{30}, що є суттєвим.

Список літератури:

1. Post-quantum cryptography: CSRC, CSRC. Available at: https://csrc.nist.gov/projects/post-quantum-cryptography (Accessed: 21 July 2024).
2. O. Regev. The Learning with Errors Problem. Available: https://cims.nyu.edu/~regev/papers/lwesurvey.pdf
3. J. Hoffstein, J. Pipher, and J. Silverman. NTRU: A Ring-Based Public Key Cryptosystem. Available: https://www.ntru.org/f/hps98.pdf
4. C. Gentry, C. Peikert, and V. Vaikuntanathan. Trapdoors for Hard Lattices and New Cryptographic Constructions. ePrint IACR, 2007. https://eprint.iacr.org/2007/432
5. C. Schnorr. Lattice Reduction by Random Sampling and Birthday Methods, 2003. Accessed: Jul. 21, 2024. [Online]. Available: https://d-nb.info/1153616645/34
6. M. Albrecht and L. Ducas. LATTICE ATTACKS ON NTRU AND LWE: A HISTORY OF REFINEMENTS. Available: https://eprint.iacr.org/2021/799.pdf
7. Y. Chen and P. Nguyen. BKZ 2.0: Better Lattice Security Estimates. Accessed: Jul. 21, 2024. [Online]. Available: https://www.iacr.org/archive/asiacrypt2011/70730001/70730001.pdf
8. S. Bai, D. Stehlé, and W. Wen. Measuring, simulating and exploiting the head concavity phenomenon in BKZ // Cryptology ePrint Archive (eprint.iacr.org), 2018. https://eprint.iacr.org/2018/856 (accessed Jul. 21, 2024).
9. Z. Zhao and G. Xu. On the Measurement and Simulation of the BKZ Behavior for q-ary Lattices // Lecture notes in computer science, pp. 463–482, Jan. 2023, doi: https://doi.org/10.1007/978-3-031-26553-2_25.
10. I. D. Gorbenko, O. G. Kachko, Y. I. Gorbenko, I. V. Stelnik, S. O. Kandy, and M. V. Yesina. METHODS OF

BUILDING GENERAL PARAMETERS AND KEYS FOR NTRU PRIME UKRAINE OF 5TH – 7TH LEVELS OF STABILITY. PRODUCT FORM // Telecommunications and radio engineering, vol. 78, no. 7, pp. 579–594, Jan. 2019, doi: https://doi.org/10.1615/telecomradeng.v78.i7.30.

11. E. Alkim, L. Ducas, T. Pöppelmann, and P. Schwabe. Post-quantum key exchange – a new hope // Cryptology ePrint Archive (eprint.iacr.org), 2015. https://eprint.iacr.org/2015/1092

12. ДСТУ 8961:2019. Інформаційні технології. Криптографічний захист інформації. Алгоритми асиметричного шифрування та інкапсуляції ключів. Чин. від 21.12.2019. Вид. офіц. Київ: УкрНДНЦ, 2019. 72 с.

13. C. Peikert. A Decade of Lattice Cryptography // Foundations and Trends® in Theoretical Computer Science, vol. 10, no. 4, pp. 283–424, 2016, doi: https://doi.org/10.1561/0400000074.

14. M. Albrecht, S. Bai, and L. Ducas. A subfield lattice attack on overstretched NTRU assumptions: Cryptanalysis of some FHE and Graded Encoding Schemes // Cryptology ePrint Archive (eprint.iacr.org), 2016. https://eprint.iacr.org/2016/127 (accessed Jul. 21, 2024).

15. L. Ducas and W. van Woerden. NTRU Fatigue: How Stretched is Overstretched? // Cryptology ePrint Archive (eprint.iacr.org), 2021. https://eprint.iacr.org/2021/999 (accessed Jul. 21, 2024).

16. C. van Vredendaal. Reduced memory meet-in-the-middle attack against the NTRU private key // LMS Journal of Computation and Mathematics, vol. 19, no. A, pp. 43–57, 2016, doi: https://doi.org/10.1112/s1461157016000206.

17. Q. Guo, T. Johansson, and P. Stankovski. Coded-BKW: Solving LWE Using Lattice Codes // Accessed: Jul. 21, 2024. [Online]. Available: https://www.iacr.org/archive/crypto2015/92160189/92160189.pdf

18. M. Albrecht, C. Cid, J.-C. Faugère, R. Fitzpatrick, and L. Perret. On the complexity of the Arora-Ge Algorithm against LWE On the complexity of the Arora-Ge Algorithm against LWE. 2012 // Accessed: Jul. 21, 2024. [Online]. Available: https://inria.hal.science/hal-00776434/PDF/SCC_AG_2012.pdf

19. D. Bernstein and T. Lange. Non-randomness of S-unit lattices. Accessed: Jul. 21, 2024. [Online]. Available: https://eprint.iacr.org/2021/1428.pdf

20. R. Lindner and C. Peikert, “Better Key Sizes (and Attacks) for LWE-Based Encryption, 2010 // Accessed: Jul. 21, 2024. [Online]. Available: https://eprint.iacr.org/2010/613.pdf

21. N. Alkadri, J. Buchmann, R. Bansarkhani, and J. Krämer. A Framework to Select Parameters for Lattice-Based Cryptography // Accessed: Jul. 21, 2024. [Online]. Available: https://eprint.iacr.org/2017/615.pdf

22. L. Bi, X. Lu, J. Luo, and K. Wang. Hybrid Dual and Meet-LWE Attack // Cryptology ePrint Archive (eprint.iacr.org), 2022. https://eprint.iacr.org/2022/1330 (accessed Jul. 21, 2024).

23. S. Bai, S. Miller, and W. Wen. A refined analysis of the cost for solving LWE via uSVP, 2019 // Accessed: Jul. 21, 2024. [Online]. Available: https://hal.science/hal-02886638/document

Надійшла до редколегії 29.04.2024

Відомості про авторів:

Кандій Сергій Олегович – Харківський національний університет імені В. Н. Каразіна, аспірант кафедри безпеки інформаційних систем і технологій, факультет комп’ютерних наук, АТ «Інститут Інформаційних технологій», науковий консультант; Україна; e-mail: sergeykandy@gmail.com; ORCID: https://orcid.org/0000-0003-0552-8341

Горбенко Іван Дмитрович – д-р техн. наук, професор, Харківський національний університет імені В.Н. Каразіна, професор кафедри безпеки інформаційних систем і технологій, факультет комп’ютерних наук, АТ «Інститут інформаційних технологій», головний конструктор; Україна; e-mail: GorbenkoI@iit.kharkov.ua; ORCID: https://orcid.org/0000-0003-4616-3449

УДК 621.391:519.2 **DOI:10.30837/rt.2024.2.217.08**

О. О. КУЗНЕЦОВ, д-р техн. наук, М. О. ПОЛУЯНЕНКО, канд. техн. наук,
Д. І. ПРОКОПОВИЧ-ТКАЧЕНКО, канд. техн. наук, Є. В. КОТУХ, канд. техн. наук,
В. О. ЛЮБЧАК, канд. фіз.-мат. наук

МОДИФІКОВАНІ ГЕНЕТИЧНІ АЛГОРИТМИ ДЛЯ ГЕНЕРАЦІЇ S-BOXES З ВИСОКОЮ НЕЛІНІЙНІСТЮ

Вступ

S-boxes (Substitution boxes, або блоки заміни) – це фундаментальні елементи криптографічних алгоритмів, які відіграють ключову роль у забезпеченні безпеки та стійкості систем шифрування [1 – 3]. Ці таблиці використовуються для заміни блоків даних на інші блоки, тим самим ускладнюючи аналіз і розшифрування закодованої інформації. Зазвичай, S-boxes приймають на вхід кілька бітів і видають на вихід кілька бітів, перетворюючи вхідні дані у вихідні через складні нелінійні функції [4 – 6].

Генерація S-boxes є критично важливим завданням у криптографії, оскільки безпека більшості сучасних криптографічних алгоритмів, таких як AES, DES, Blowfish, Twofish та багатьох інших, значною мірою залежить від стійкості та надійності їх S-boxes [3, 7, 8]. Ці блоки заміни забезпечують необхідну нелінійність та дифузію, що захищає дані від різних типів атак, включаючи лінійний та диференціальний криптоаналіз.

Створення надійних S-boxes є складним завданням, оскільки атакуючі постійно вдосконалюють свої методи криптоаналізу, намагаючись знайти слабкі місця у криптографічних системах. Тому S-boxes повинні мати специфічні властивості, які забезпечують їх стійкість до різних атак [9 – 11]:

1. **Нелінійність**: S-box повинен бути високонелінійним, щоб зміна одного біта у вхідному блоці призводила до зміни більше одного біта у вихідному блоці. Це ускладнює застосування лінійного криптоаналізу.
2. **Рівномірність**: Вихідні значення S-box повинні бути розподілені рівномірно, що забезпечує рівноймовірність кожного можливого виходу та захист від диференціального криптоаналізу.
3. **Стійкість до атак**: S-box повинен бути стійким до різних атак, таких як лінійний та диференціальний криптоаналіз, а також до методів зворотного аналізу.
4. **Інваріантність**: S-box повинен бути інваріантним до перестановки вхідних або вихідних бітів, що підвищує його стійкість до атак, заснованих на перестановці бітів.
5. **Різноманітність**: S-box повинен бути унікальним і відрізнятися від інших S-boxes, щоб запобігти атакам з використанням попередньо обчислених таблиць.

Забезпечення всіх цих властивостей при генерації S-boxes є важливою задачею для забезпечення надійності та безпеки криптографічних алгоритмів. Існує кілька методів генерації S-boxes, включаючи випадкову генерацію, методи лінійних перетворень, комбінаційну оптимізацію та генетичні алгоритми. Кожен з цих методів має свої переваги та недоліки, і вибір методу залежить від конкретних вимог до системи [12 – 14].

1. **Випадкова генерація**: Найпростіший метод, який, однак, може не забезпечити достатньої стійкості та безпеки.
2. **Метод лінійних перетворень**: Забезпечує хорошу стійкість до лінійного криптоаналізу, але може бути вразливим до інших атак.
3. **Комбінаційна оптимізація**: Дає можливість створити S-boxes з високою стійкістю до різних атак, але є обчислювально складним.
4. **Генетичний алгоритм**: Забезпечує ефективну генерацію S-boxes зі значною стійкістю до атак, при цьому є обчислювально ефективним.

Застосування генетичних алгоритмів для генерації S-boxes є перспективним напрямком досліджень, оскільки вони дозволяють досягти високого рівня стійкості при помірній обчислювальній складності. У цій статті розглядається генетичний алгоритм для генерації S-boxes. Ми досліджуємо ефективність базової версії алгоритму, демонструємо високу складність генерації нелінійних підстановок, а також пропонуємо новий алгоритм селекції, що підвищує ефективність генерації. Результати експериментів показують, що запропонований метод значно підвищує ефективність і надійність S-boxes.

2. Генетичні алгоритми для генерації S-box

2.1. Загальні відомості про генетичні алгоритми

Генетичні алгоритми (ГА) є потужним методом оптимізації, який базується на принципах природного відбору та генетики. Вперше запропоновані Джоном Холландом у 1970-х роках [15, 16], ці алгоритми моделюють процес еволюції популяцій організмів з метою пошуку оптимальних або близько оптимальних розв'язків складних задач.

Основні етапи генетичних алгоритмів включають [16, 17]:

1. **Ініціалізація популяції**: Генетичний алгоритм починається зі створення початкової популяції індивідів (можливих розв'язків). Кожен індивід кодується у вигляді хромосоми, яка може бути бінарною стрічкою, вектором чисел чи іншою структурою даних.
2. **Оцінка пристосованості (фітнес-функція)**: Кожен індивід оцінюється за допомогою фітнес-функції, яка вимірює якість або пристосованість індивіда до вирішення задачі.
3. **Відбір**: На основі фітнес-значень відбираються індивіди, які братимуть участь у створенні нової популяції. Частіше відбираються індивіди з вищою пристосованістю.
4. **Кросовер (схрещування)**: Відбрані індивіди (батьки) схрещуються між собою для утворення нових індивідів (нащадків). Кросовер дозволяє комбінувати генетичну інформацію батьків і створювати нові розв'язки.
5. **Мутація**: Застосовується випадкова зміна генів у хромосомах нащадків для підтримки генетичної різноманітності популяції і запобігання передчасній збіжності до локальних оптимумів.
6. **Заміна**: Старе покоління індивідів замінюється новим поколінням, і процес повторюється, поки не буде досягнуто критерію зупинки, наприклад, максимальна кількість поколінь або задовільний рівень фітнесу.

2.2. Застосування генетичних алгоритмів для генерації S-box

Генерація S-boxes за допомогою генетичних алгоритмів є перспективним підходом завдяки їх здатності знаходити складні, нелінійні і стійкі до атак структури. Основні етапи застосування ГА для генерації S-boxes включають [11, 18, 19]:

1. **Ініціалізація популяції S-boxes**: Початкова популяція складається з випадково згенерованих S-boxes. Кожна S-box представлена у вигляді хромосоми, де кожен ген відповідає певному значенню таблиці заміни.
2. **Оцінка пристосованості S-boxes**: Для кожної S-box обчислюється фітнес-функція, яка враховує кілька критеріїв, таких як нелінійність, рівномірність розподілу, стійкість до диференціального та лінійного криптоаналізу. Ця функція може бути комбінацією різних метричних показників.
3. **Відбір найбільш пристосованих S-boxes**: Використовуються методи відбору, такі як турнірний відбір або відбір за пропорцією пристосованості, для вибору батьківських S-boxes для наступних поколінь.
4. **Кросовер і мутація**: Відібрані S-boxes піддаються операціям кросоверу і мутації для утворення нових S-boxes. Кросовер може здійснюватися шляхом обміну частин хромосом між батьківськими S-boxes, а мутація – шляхом випадкових змін окремих значень у таблиці заміни.

5. **Заміна і еволюція**: Нові покоління S-boxes замінюють старі, і процес повторюється, поки не буде досягнуто оптимальних значень фітнес-функції або максимальна кількість поколінь.

Застосування генетичних алгоритмів для генерації S-boxes дозволяє ефективно знаходити рішення, що відповідають вимогам до стійкості та безпеки криптографічних алгоритмів. Генетичні алгоритми забезпечують широкий простір пошуку та гнучкість в налаштуванні параметрів, що дозволяє оптимізувати процес генерації S-boxes для конкретних криптографічних потреб.

3. Методологія досліджень та результати

3.1. Методологія досліджень

Для дослідження ефективності генетичних алгоритмів у генерації S-boxes була обрана цільова функція WHS (Walsh–Hadamard Spectrum), запропонована Кларком у роботі [20]. Ця функція має вигляд

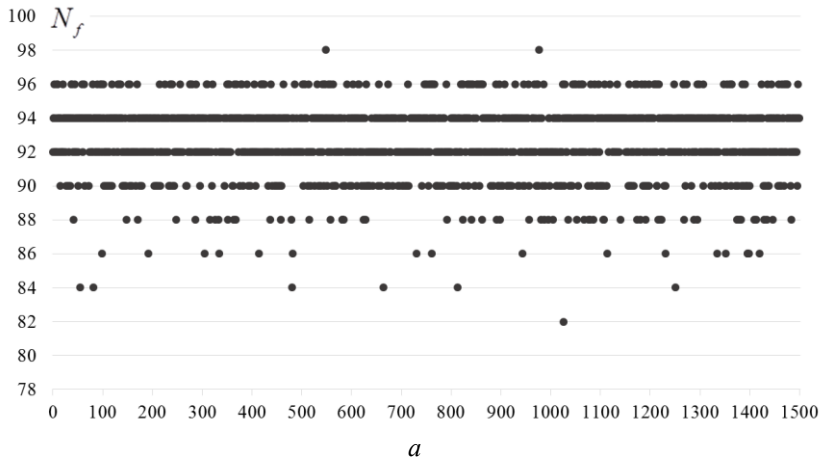
$$WHS = \sum_{b=1}^{255}\sum_{i=0}^{255} \left| |WHT[b,i]| - X \right|^R ,$$

де WHT (англ. Walsh–Hadamard transform) – спектральні коефіцієнти Уолша–Адамара; i – цикл за всіма компонентними функціями та їх лінійними комбінаціями; b – цикл за всіма лінійними функціями; X і R – параметри з дійсними значеннями.

В якості оптимальних параметрів функції *WHS* було обрано $R = 12$ та $X = 0$. При оптимальних параметрах та використанні алгоритму пошуку сходження на пагорб з ймовірністю близької до 99 % може бути сформований бієктивний S-блок з нелінійністю 104 [21, 22].

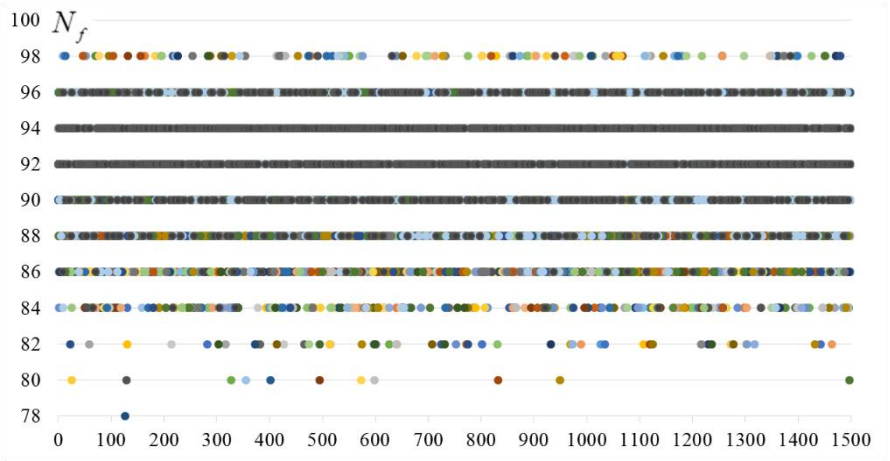
Генетичний алгоритм передбачає вибір найкращого результату серед нащадків. У початку алгоритму пошуку дуже швидко можливо знайти покращення результату при випадковому виборі мутацій стану S-блоку. Але чим кращий поточний стан S-блоку тім складніше знайти покращення та майже завжди всі зміни S-блоку будуть мати гірше значення.

Ми провели ряд запусків генетичного алгоритму з цільовою функцією *WHS* з найкращими її параметрами $R = 12$ та $X = 0$, кількість ітерації була обмежена значенням у 150 000, кожні 100 ітерацій фіксувався поточний стан: значення цільової функції та нелінійність. В якості прикладу на рис. 1, *а* наведено отримані послідовність зміни нелінійності для одного запуску алгоритму пошуку. На рис. 1, *б* наведено результат 50 таких запусків. Для кожного запуску ми використовували п’ять різних мутацій та кількість екземплярів у популяції дорівнювало п’яти. Як бачимо, жодного разу нелінійність не досягла хоча б значення у $N_f = 100$.



а

Рис. 1. Зміна нелінійності через кожні 100 ітерацій (всього 150 000 ітерацій):
а – результат одного запуску, *б* – результат за 50 запусками



б

Рис. 1. (Продовження)

На рис. 2 наведено гістограму розподілу частоті нелінійності для тих же результатів. На рис. 3 наведено частоту нелінійності для випадково сформованого S-блоку отриманих на вибірки з 10^8 формувань.

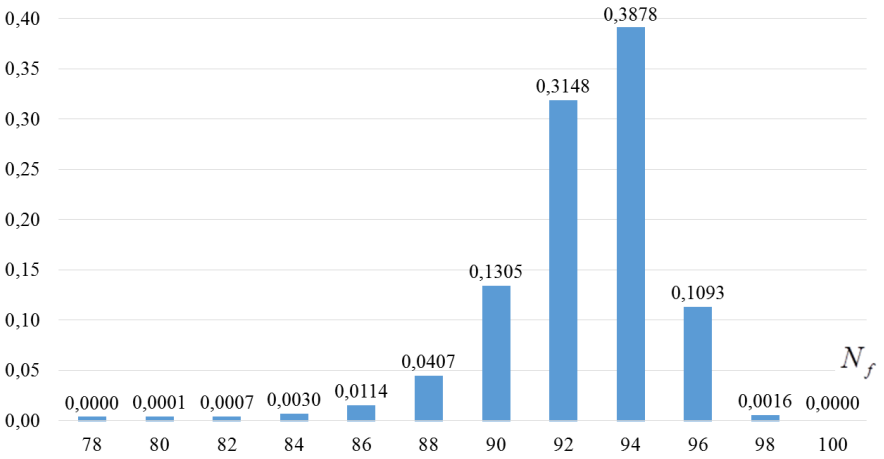


Рис. 2. Гістограма розподілу нелінійності за резльутатами 75 000 замеров впродовж 50 окремих запусків алгоритму пошуку

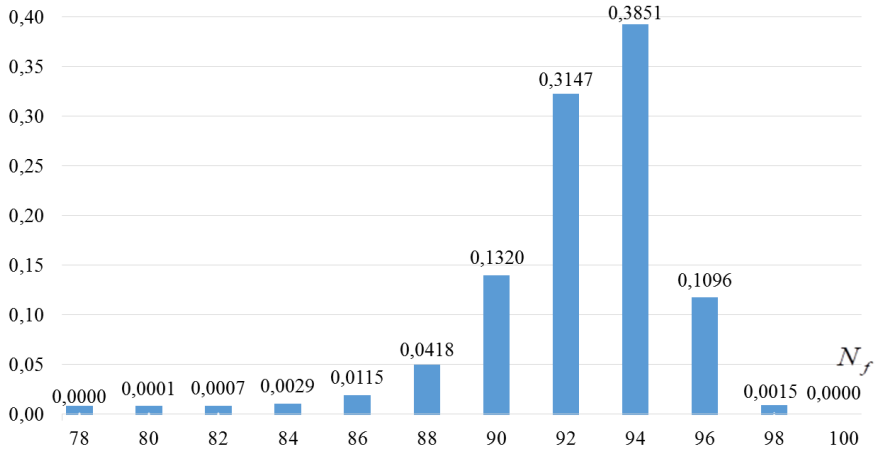


Рис. 3. Розподіл кількості сформованих випадковим чином S-блоків в залежності від їх нелінійності (N_f) при 10^8 випробуваннях

Як бачимо, розподіли майже ідентичні, що вказує на відсутність прогресу покращення пошуку генетичним алгоритмом у його «класичному» виді та обраних параметрів пошуку. Тому ми дещо змінили алгоритм пошуку, додавши зберігання до популяції не лише нащадків, а ще й поточну популяцію. Відбраковування у популяції зайвої кількості S-блоків будемо проводити за принципом відбору найкращих за значенням цільової функції. Данна модифікація алгоритму пошуку за функціональністю стала походити на методи Hill Climbing.

3.2. Алгоритм селекції

Основна ідея методу селекції полягає у розгляді деякої популяції S-блоків, яка формується випадковим чином. Проведення мутації у цієї популяції. Під мутацією будемо розуміти зміну містами двох випадково обраних (неоднакових) позицій у S-блоці; оцінку якості отриманих нових S-блоків за допомогою деякої цільової функції; ранжирування S-блоків зі старої популяції та нових S-блоків; створення нової популяції найкращих екземплярів S-блоків згідно зі створеним ранжируваним списком S-блоків (обираючи до нової популяції найкращих представників); повторення наведених дії до тих пір, поки не буде знайдено цільовий S-блок або не виконано інші критерії зупинки алгоритму.

Псевдокод алгоритму селекції наведено на рис. 4.

Вхід: S_{pop}, K_{iter}, K_{pop}
For ($t = 0$ to $t < K_{\text{iter}}$, $t = t + 1$){
    $S_{\text{pop}} = \text{basic\_selection}\left[S_{\text{pop}} \right]$;
    For ($p = 0$ to $p < K_{pop}$, $p = p + 1$){
        $S \leftarrow S_{\text{pop}}[p]$;
        For ($k = 0$ to $k < K_{\text{mut}}$, $k = k + 1$){
            $S' \leftarrow S$;
            $i \leftarrow random[0...255]$;
            $j \leftarrow random[0...255]$;
            $swap(S'[i], S'[j])$;
            $N_f, F_c \leftarrow f(S')$;
            If ($N_f \le 104$) Return S';
            $S_{\text{pop}} = S_{\text{pop}} + S'$;
        }
    }
}
Return 0.

Рис. 4. Псевдокод алгоритму селекції

На вхід алгоритм селекції отримує:

- K_{pop} – кількість екземплярів у популяції – кількість найкращий S-блоків яку зберігаємо у популяції;
- S_{pop} – випадковим чином сформована популяція бієктивних S-блоків. В нашому випадку, для початкового формування бієктивного S-блоку, ми використовували алгоритм Фішера – Йетса;
- K_{iter} – максимальна кількість ітерацій виконання алгоритму селекції. Є одним з критеріїв зупинки алгоритму, якщо він не в змозі подолати локальний мінімум;

- K_{mut} – кількість мутацій – кількість нових екземплярів (S-блоків), отриманих з кожного представника (S-блоку) з поточної популяції.

Ітеративно виконуємо крок алгоритму селекції. Виконуємо не більше $K_{\text{iter}} = 150000$ ітерацій. На початку кожної ітерації виконуємо селекцію у популяції за допомогою функції basic_selection:

basic_selection – функція, яка виконує ранжирування списку S-блоку за значенням цільової функції та значенням нелінійності. До топу заносяться S-блоки, які мають більш високу нелінійність, S-блоки, які мають однакову нелінійність сортуються за значенням цільової функції: чим менше значення – тим ближче до топу поміщується S-блок. На виході функції залишається S-блоки з топу у кількості K_{pop} штук.

З кожним окремим екземпляром популяції з S_{pop} виконуємо алгоритм мутації K_{mut} раз. В якості алгоритму мутації нами обрана заміна містами двох випадково вибраних (неоднакових) елементів у S-блоці. Одночасно обчислюєтеся її цільова функція (F_c) та перевіряється значення нелінійності (N_f). Якщо нелінійність відповідає бажаному значенню, то завершуємо алгоритм селекції та повертаємо знайдений S-блок. У іншому випадку додаємо до популяції новий створений екземпляр. Наприкінці роботи кроку ітерації маємо популяцію розміром $K_{\text{pop}} \times (K_{\text{mut}} + 1)$, яку на наступній ітерації, за допомогою функції basic_selection, зменшимо до розміру K_{pop}.

3.3. Отримані результати

Враховуючи, що обчислення значення цільової функції є сама затратна (з точки зору процесорного часу) операція, складність обчислювання всього алгоритму пошуку можна вважати пропорційною кількості викликів обчислення цільової функції. Тобто кількості S-блоків, які було сформовано та перевірено. Позначимо таку кількість як K_{Sbox}.

Для прискорення роботи алгоритму було проведено паралельне обчислення нової популяції у $N_{thread} = 8$ потоках всередині кожної ітерації.

Алгоритм селекції виконувався при різних значеннях $K_{\text{pop}} = [1,21]$ (з кроком 2) та $K_{\text{mut}} = [1,31]$ (з кроком 3). Для кожного значення виконувалось по 100 запусків алгоритму пошуку.

Усереднене значення складності пошуку наведено у таблиці 1 та візуалізовано на рис. 5. Найкращі параметри, з точки зору меншої кількості обчислень цільової функції, були при $K_{\text{pop}} = 1$. Усереднене значення K_{Sbox} коливається у діапазоні від 49 277 до 58 213. Слід зазначити, що при кожному випробуванні мале місце дуже велике коливання від середнього значення (приклад для кращого значення у $K_{Sbox} = 49277$ при $K_{\text{pop}} = 1$ та $K_{\text{mut}} = 7$, наведено на рис. 6). Середньоквадратичне відхилення для $K_{\text{pop}} = 1$ складало 22 500 а для $K_{\text{pop}} = 21$ – 44 500.

Враховуючи найкращі результати при малих значеннях K_{pop}, для K_{pop}=1 та 2, було виконано ще серію зі 100 запусків для $K_{\text{mut}} = [1,30]$ (з кроком 1). Результати наведено на рис. 7. При цьому середньостатистична кількість K_{Sbox} становить близько 53 000 для $K_{\text{pop}} = 1$.

Майже завжди алгоритм пошуку знаходив S-блок з нелінійністю $N_f = 104$. Лише у 2 % для випадку $K_{\text{pop}} = 1$ та $K_{\text{mut}} = 1$ алгоритм зупинився у локальному мінімумі, не досягнувши мети пошуку, та у 1 % – для значень $K_{\text{pop}} = 1$ та $K_{\text{mut}} = 6,8,11$.

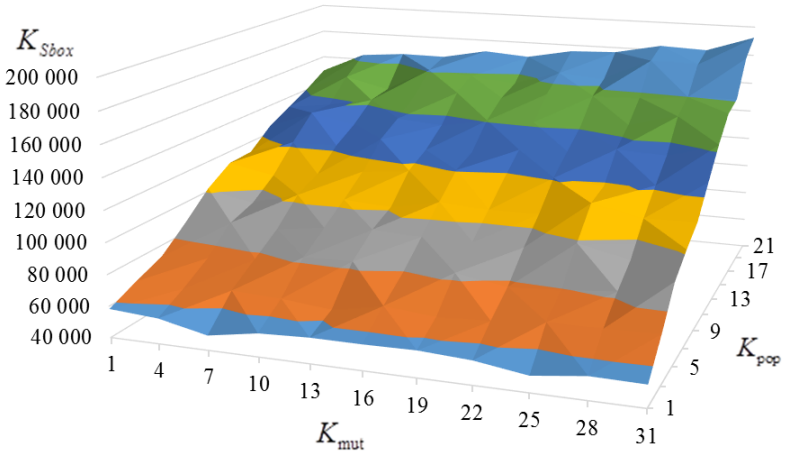


Рис. 5. Усереднене значення кількісті сформованих S-блоків (K_{Sbox}) до знаходження S-блоку з $N_f = 104$

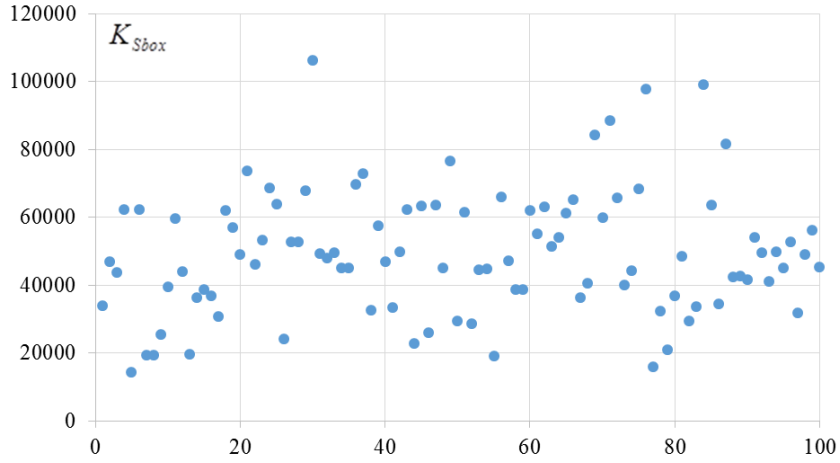


Рис. 6. Кількість S-блоків, які було сформовано та перевірено (K_{Sbox}) до знаходження $N_f = 104$ при $K_{pop} = 1$ та $K_{mut} = 7$

Таблиця 1

Усереднене значення кількості сформованих S-блоків (K_{Sbox}) до знаходження S-блоку з $N_f = 104$

	K_{pop}										
K_{mut}	1	3	5	7	9	11	13	15	17	19	21
1	58 213	65 942	72 830	86 642	101 726	111 990	112 718	125 113	132 806	140 336	149 339
4	56 067	64 863	75 069	89 598	94 726	105 925	122 364	137 003	136 740	151 874	163 291
7	49 277	67 198	77 848	88 353	103 154	109 618	122 382	130 901	142 463	144 601	165 918
10	54 636	65 723	82 198	92 542	102 797	114 163	129 411	137 442	147 416	161 020	165 672
13	56 042	62 660	83 216	94 538	101 073	117 611	124 466	135 244	152 048	158 696	171 756
16	56 010	68 711	79 645	93 134	107 371	120 567	125 274	140 817	150 494	155 049	169 462
19	56 532	65 910	82 883	92 911	105 144	117 877	129 718	142 017	155 463	164 902	175 531
22	54 775	67 236	77 663	92 874	105 559	120 992	131 029	140 772	156 224	162 808	176 669
25	50 066	70 394	79 596	98 967	115 462	118 406	135 294	144 708	157 321	177 621	183 087
28	54 203	70 453	82 200	91 841	108 783	121 683	133 665	152 984	159 887	176 751	181 781
31	53 709	71 581	91 827	101 536	109 625	126 616	143 233	156 987	160 573	183 069	192 493

Таблиця 2

Усереднений час (секунди) до знаходження S-блоку з $N_f = 104$

	K_{pop}										
K_{mut}	1	3	5	7	9	11	13	15	17	19	21
1	64	23	17	16	15	18	16	20	19	19	42
4	66	28	17	11	9	14	14	20	18	19	30
7	57	32	21	10	9	14	13	18	18	17	28
10	42	33	25	10	9	15	14	19	19	19	24
13	44	30	26	10	8	15	13	19	20	19	22
16	44	32	24	10	9	15	14	19	19	18	20
19	56	35	24	13	8	15	14	19	19	19	19
22	84	34	21	21	9	15	14	19	19	19	18
25	46	26	18	17	16	15	14	19	20	21	18
28	62	24	19	15	14	15	14	21	20	21	18
31	69	30	20	15	14	16	15	21	20	21	18

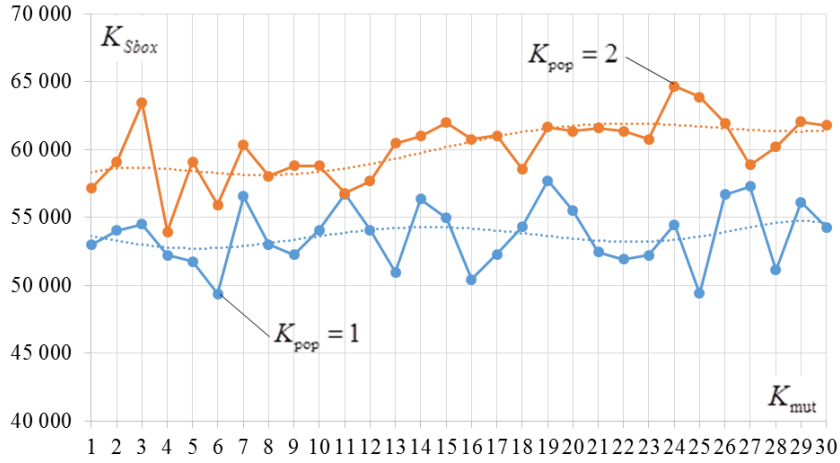


Рис. 7. Усереднене значення кількості сформованих S-блоків (K_{Sbox}) до знаходження S-блоку з $N_f = 104$ для K_{pop}=1 та 2

Таким чином, найменша кількість K_{Sbox} для алгоритму селекції становить близько 53 000 при $K_{pop} = 1$ та $K_{mut} = 4-9$. З ймовірністю 99 % під час пошуку буде знайдено S-блок з нелінійністю $N_f = 104$.

4. Обговорення результатів дослідження

4.1. Обговорення результатів

Отримані результати показують, що модифікований генетичний алгоритм з функцією WHS та оптимальними параметрами R=12 та X=0 є ефективним засобом для генерації S-boxes з високою нелінійністю. Використання селекції та додаткового зберігання поточної популяції дозволило значно покращити ефективність алгоритму порівняно з класичним генетичним алгоритмом.

Наші експерименти показали, що найкращі результати досягаються при використанні популяції з одного екземпляра та кількості мутацій у діапазоні від 4 до 9. У середньому, алгоритм вимагав перевірки близько 53 000 S-блоків для знаходження S-блоку з нелінійністю $N_f = 104$. Важливо зазначити, що при малих значеннях K_{pop} спостерігалися значні коливання від середнього значення, що свідчить про високу варіативність процесу пошуку.

Модифікований алгоритм селекції показав свою перевагу перед класичним генетичним алгоритмом. У більшості випадків він досягав мети з нелінійністю $N_f = 104$ з ймовірністю 99 %. Це підтверджує ефективність обраної стратегії селекції та підходу до мутацій, що забезпечують високий рівень стійкості до атак.

Розподіл частот нелінійності, отриманий у результаті наших експериментів, показав близьку відповідність до теоретичних очікувань. Це свідчить про правильність вибраних параметрів та налаштувань алгоритму. Додатково було виявлено, що середньоквадратичне відхилення результатів зменшується з ростом кількості мутацій, що також впливає на стабільність пошуку.

4.2. Порівняння з іншими методами

У порівнянні з іншими методами генерації S-boxes, такими як випадкова генерація або методи лінійних перетворень, наш підхід має кілька ключових переваг. Генетичний алгоритм з модифікованою селекцією дозволяє ефективно знаходити S-boxes з високими показниками нелінійності та стійкості до криптоаналізу. Крім того, він забезпечує гнучкість у налаштуванні параметрів, що дозволяє адаптувати алгоритм під конкретні вимоги криптографічних систем.

5. Висновки

У дослідженні розглянуто використання генетичних алгоритмів для генерації S-boxes з високою нелінійністю. Запропонований підхід базувався на використанні цільової функції WHS з оптимальними параметрами R=12 та X=0, що дозволило досягти високої ефективності у генерації бієктивних S-блоків.

Результати дослідження показали, що модифікований генетичний алгоритм з додатковим зберіганням поточної популяції та використанням селекції значно покращив ефективність пошуку. Найкращі результати були досягнуті при $K_{\text{pop}} = 1$ та $K_{\text{mut}} = 4 - 9$, де середня кількість перевірених S-блоків становила близько 53 000.

Запропонований алгоритм показав високу стабільність та ефективність у генерації S-boxes з нелінійністю $N_f = 104$, що підтверджується 99 % ймовірністю досягнення мети. Ці результати свідчать про перспективність використання генетичних алгоритмів у криптографічних застосуваннях, де потрібна висока стійкість до атак.

У майбутньому планується подальше вдосконалення алгоритму шляхом дослідження інших методів мутацій та селекції, а також оптимізація параметрів для досягнення ще кращих результатів. Крім того, можливе використання розподілених обчислень для подальшого прискорення процесу генерації S-boxes.

Список літератури:

1. Mishra N., Hafizul Islam S., Zeadally S. A survey on security and cryptographic perspective of Industrial-Internet-of-Things // Internet of Things. 2024. Vol. 25. P. 101037.
2. Urooj S. et al. Cryptographic Data Security for Reliable Wireless Sensor Network // Alexandria Engineering Journal. 2023. Vol. 72. P. 37–50.
3. Tiwari A. Chapter 14 – Cryptography in blockchain // Distributed Computing to Blockchain ; ed. Pandey R., Goundar S., Fatima S. Academic Press, 2023. P. 251–265.
4. A S C., S P M., R K P. Implementation of S-box for lightweight block cipher // 2023 3rd International Conference on Intelligent Technologies (CONIT). 2023. P. 1–4.
5. R M., V N.K. Optimized Implementation of S-box and Inverse S-box for PRESENT Lightweight Block Cipher // 2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN). 2023. P. 1–5.
6. Teja P.R., Sasamal T.N. Implementation of Efficient Serial Architecture for Prince Block Cipher with Enhanced Security // 2023 International Conference on System, Computation, Automation and Networking (ICSCAN). 2023. P. 1–6.
7. Grami A. Chapter 11 – Cryptography // Discrete Mathematics ; ed. Grami A. Academic Press, 2023. P. 197–210.

8. Milanič M., Servatius B., Servatius H. Chapter 8 – Codes and cyphers // Discrete Mathematics With Logic / ed. Milanič M., Servatius B., Servatius H. Academic Press, 2024. P. 163–179.

9. McLaughlin J. Applications of search techniques to cryptanalysis and the construction of cipher components: phd. University of York, 2012.

10.Álvarez-Cubero J. Vector Boolean Functions: applications in symmetric cryptography. 2015.

11.Burnett L.D. Heuristic Optimization of Boolean Functions and Substitution Boxes for Cryptography: phd. Queensland University of Technology, 2005.

12.Clark A.J. Optimisation heuristics for cryptology: phd. Queensland University of Technology, 1998.

13.Fuller J.E. Analysis of affine equivalent boolean functions for cryptography: phd. Queensland University of Technology, 2003.

14.Carlet C., Ding C. Nonlinearities of S-boxes // Finite Fields and Their Applications. 2007. Vol. 13, № 1. P. 121–135.

15.Ghosh A., Das S., Saha B. Chapter 6 - Nature-inspired optimization algorithms // Artificial Intelligence in Textile Engineering / ed. Ghosh A., Das S., Saha B. Woodhead Publishing, 2024. P. 171–231.

16.Tsai C.-W., Chiang M.-C. Chapter Seven - Genetic algorithm // Handbook of Metaheuristic Algorithms ; ed. Tsai C.-W., Chiang M.-C. Academic Press, 2023. P. 111–138.

17.Tsai C.-W., Chiang M.-C. Chapter Fifteen – Hybrid metaheuristic and hyperheuristic algorithms // Handbook of Metaheuristic Algorithms ; ed. Tsai C.-W., Chiang M.-C. Academic Press, 2023. P. 321–350.

18.Tesar P. A New Method for Generating High Non-linearity S-Boxes. Společnost pro radioelektronické inženýrství, 2010.

19.Ivanov G., Nikolov N., Nikova S. Reversed genetic algorithms for generation of bijective s-boxes with good cryptographic properties // Cryptogr. Commun. 2016. Vol. 8, № 2. P. 247–276.

20.Clark J.A., Jacob J.L., Stepney S. The design of s-boxes by simulated annealing // Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No.04TH8753). 2004. Vol. 2. P. 1533–1537 Vol.2.

21.Kuznetsov A. et al. WHS Cost Function for Generating S-boxes // IEEE Int. Conf. Probl. Infocommunications, Sci. Technol., PIC S T – Proc. Institute of Electrical and Electronics Engineers Inc., 2021. P. 434–438.

22.Kuznetsov A. et al. Opportunities to minimize hardware and software costs for implementing boolean functions in stream ciphers // Int. J. Comput. Research Institute of Intelligent Computer Systems, 2019. Vol. 18, № 4. P. 443–452.

Надійшла до редколегії 08.06.2024

Відомості про авторів:

Кузнецов Олександр Олександрович – д-р техн. наук, професор, професор кафедри безпеки інформаційних систем і технологій; Харківський національний університет імені В. Н. Каразіна; Харків, Україна; e-mail: kuznetsov@karazin.ua, n.poluyanenko@karazin.ua; ORCID: https://orcid.org/0000-0003-2331-6326

Полуяненко Микола Олександрович – канд. техн. наук, доцент, доцент кафедри безпеки інформаційних систем і технологій; Харківський національний університет імені В. Н. Каразіна; Харків, Україна; e-mail: n.poluyanenko@karazin.ua; ORCID: https://orcid.org/0000-0001-9386-2547

Прокопович-Ткаченко Дмитро Ігорович – канд. техн. наук, доцент, доцент кафедри кібербезпеки та інформційних технологій; Університет митної справи та фінансів; Дніпро, Україна; e-mail: omega2417@gmail.com; ORCID: https://orcid.org/0000-0002-6590-3898

Котух Євген Володимирович – канд. техн. наук, професор кафедри кібербезпеки; Національний технічний університет «Дніпровська політехніка»; Дніпро, Україна; e-mail: yevgenkotukh@gmail.com; ORCID: https://orcid.org/0000-0003-4997-620X

Любчак Володимир Олександрович – канд. фіз.-мат. наук, доцент, завідувач кафедри кібербезпеки; Сумський державний університет; e-mail: v.liubchak@dcs.sumdu.edu.ua; ORCID: https://orcid.org/0000-0002-7335-6716

MEANS OF TELECOMMUNICATIONS
ЗАСОБИ ТЕЛЕКОМУНІКАЦІЙ

УДК 621.396 **DOI:10.30837/rt.2024.2.217.09**

O.J. KADATSKAYA, PhD, C.O. SABUROVA

SUPPORT OF RESOURSES REDISTRIBUTION IN NB-IOT LTE NETWORKS

Introduction

The world has begun large-scale deployment of 5G generation wireless broadband access networks, developed by 3GPP consortium and called New Radio (NR). However, the rapid development of the concept of IoT has become the reason for the need to provide wireless connectivity a huge number of devices that are not tied to a specific subscriber but are part of the infrastructure. Within the 5G NR standard for such devices, a new type of service is provided, which called massive machine type communication (mMTC) and is focused on optimizing the use of network resources to support a large number of stable connections per unit of area. However, due to the requirements of mobility, extended coverage area, security, diverse QoS, etc., a large percentage of MTCDs will need to connect directly to cellular networks [1].

Machine-to-machine (M2M) is expected to significantly increase in future wireless networks, M2M communications enable direct communication between multiple devices. In recent years, the number of MTC devices has grown tremendously. As one of the most promising technologies for fifth generation (5G) mobile networks, MTC will enable innumerable applications in areas such as smart homes, healthcare, automotive communications, and intelligent cities. For both MTC are primarily characterized by a high device density, a low data rate, an acceptable level of delay tolerance, and high connection/communication frequency.

Improving the transmission link performance

Narrowband Internet of Things (NB-IoT) is an LPWAN protocol standardized by 3GPP that enables a wide range of new IoT devices and services connected to the cellular network [2].

The Internet of Things (IoT) is a global infrastructure for the information society that provides the ability to provide heavier services by connecting (physical and virtual) things to each other based on existing interoperable information and communication technologies. NB-IoT designed for stationary devices with low data transfer and low consumption. Data collection has played a key role throughout the era of the Internet of Things. More and more devices are interconnected, and wireless applications have become the preferred networking solution. A provider of IoT solutions continues to develop a wide range of wireless sensor devices for a variety of applications in order to provide customers with the latest solutions to complement their IoT application systems.

There are obviously many parts in any IoT device: sensors, actuators, boards, antennas, chips, micro-electro-mechanical systems and so forth. The data, which are a result of the sensing and converting of any given state or change of state in temperature, presence of gases, location and so forth usually, go from the sensor hub or IoT gateway to the cloud or a datacenter. However, given the described movement to the edge and the increasing functions of IoT gateways and IoT platforms a lot of IoT data processing and preparation (including analysis) can happen close to the devices (the edge) or in the mentioned gateways and platforms. At the same time, one of the main problems is the problem of traffic control on the radio interface in order to ensure the specified quality standards (QoS) for each service provided to the majority of subscribers, in particular, for those who are in corporate networks.

Telecommunications and information technology are now quickly becoming outdated due to the constant updating of technological solutions. However, thanks to WISE-4000 cloud access, data can be transferred directly to the cloud without using a gateway. Traditional automation architecture and basic data acquisition are no longer sufficient to collect various types of data for various IoT

applications, so companies are developing wireless sensor nodes (e.g. Advantech WISE-4000 (WSN), WISE-4000) based on the latest IoT concepts and technologies. and are a ready-to-use cloud reading and writing tool.

The REST communication approach can take advantage of not having to use a lot of bandwidth when transferring data. RESTful web APIs in JSON format makes it easy to integrate data into IoT services and optimize them for use over the Internet. In addition, REST supports HTTPS or TLS, which increase security when publishing or receiving data between devices and the cloud. In addition, it also allows end devices to actively public.

Experts believe that NB-IoT technology will gain popularity among operators, because its maintenance and operation will cost them less than today's advanced LTE and GSM networks [3], this is due to its characteristics.

Massive NB-IoT modules that try to request the radio channel resources at the same time for uplink data transmission may suffer from random access preamble collision. This is caused by several factors such as detection inaccuracy that may not satisfy the detection threshold, the high probability of false alarm, etc. Several works have proposed random access preamble detection algorithms (i.e., random access with differential barring etc.) and others have developed mathematical models to characterize the preamble transmissions in order to improve the NPRACH success rate and better time-of-arrival estimation and other NPRACH performance improvements. However, it is still unclear which scheme is effective for massive deployment, since most of the proposed schemes do not consider the heterogeneous network architecture, channel estimation impairments, or realistic channel conditions [3,4].

An IoT module is a small electronic device embedded in objects, machines, and things connected to wireless networks and sends and receives data. IoT modules and IoT terminals use a variety of wireless technologies to stay seamlessly and securely connected.

These range for:

- 5G, 4G, and 3G cellular solutions for high bandwidth applications like connected cars,
- Low-Power, Wide-Area (LPWAN) solutions such as MTC (Machine Type Communication),
- Bluetooth and LoRa are used for intelligent road systems, smart city applications, and enterprise applications.

Selecting the IoT RF module with the best features, bandwidth, and price point for each use case is an essential step toward achieving business and revenue goals.

These embedded modules dramatically shorten the development time required to complete sophisticated IoT, Industrial Internet of Things (IIoT), and automation projects. To succeed, organizations must have highly developed data and analytics practices, robust operational technologies and agile chops, as well as the ability to navigate change. Tibbo's IoT module lineup includes the WM2000 featuring an integrated Wi-Fi interface.

The problem is, most IoT solutions don't need the sizzling speed and broad bandwidth of 4G. It doesn't make sense to pay for 4G when 2G capabilities are necessary. 4G mobile network can become an LPWAN (Low Power Wide Area Networks) network with a simple software upgrade. LPWAN can connect a vast sea of IoT objects, improving safety, efficiency, and resource management by delivering on the 3C's of IoT applications:

1) Cost: LPWAN is to cut more than 50% of the cost compared to broadband LTE.

2) More than 100x lower power than broadband LTE.

3) Coverage: 5x greater coverage than broadband LTE in terms of gain.

IoT LTE modules based on optimized "categories" of LPWAN 4G cellular technologies (LTE Cat 1, Cat M, Cat NB-IoT) drive significant cost savings, efficiency, and device simplicity compared to broadband LTE. Cat 1, Cat M and Cat NB-IoT LTE modules allow highly efficient use of the current LTE spectrum requiring far less power and delivering the connectivity necessary for most IoT applications.

Also, LPWAN LTE modules are much less complicated, allowing for cost-efficient design and they offer longer-range connectivity with in-depth coverage. Choosing the right IoT wireless module will ease development, speed up time to market, and ultimately improve ROI.

Mathematical model of processing traffic intensity of macro - and microcells in NB-IoT LTE network

Programmable modules Internet of Things (IoT) are highly integrated, compact embedded devices with integrated Ethernet or Wi-Fi that serve as the foundation of own hardware solutions. IoT solution design, consideration must be given to determine the necessary features. Still, it's only possible with pervasive IoT solutions built on flexible and long-lived wireless connectivity. At the heart of it all is a tiny device called the IoT module responsible for connecting virtually anything to wireless networks. IoT Modules come with a wide range of wireless technology standards, and they provide a variety of features that can impact the success of IoT applications.

Propose to expand the capabilities of the programmable IoT module, using a mathematical model describing a of a cluster of macro- and microcells with NB-IoT LTE network services.

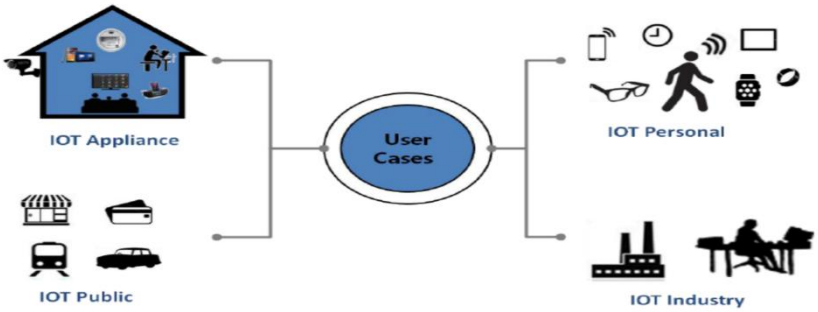


Fig. 1 Model of implementation of NB-IoT LTE services

The NB-IoT LTE standard can deployed in three versions:

– autonomous (standalone);

– on the guard-band;

– in-band.

The most common is in-band, it is widely used, where NB-IoT in-band networks are deployed by telecommunications companies Vodafone, Deutsche Telekom, Telecom Italia Mobile and others.

Mobile LTE networks involve a hierarchical organization of cells, which improves service quality and increases the efficiency of bandwidth use [6]. Will to analyze the bandwidth of a fragment of the hierarchical mobile network.

Services providers (SPs) in LTE systems are enduring many challenges in order to accommodate the rapid expansion of mobile data usage.

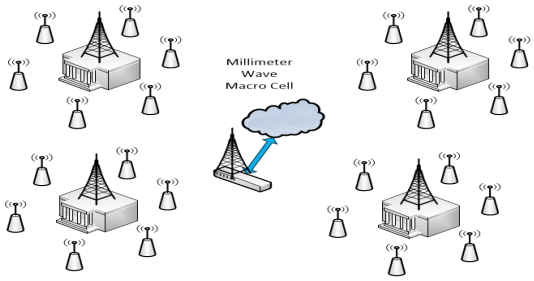


Fig. 2. LTE wo-tier cellular network topology

At the same time, Narrowband refers to NB-IoT's bandwidth of maximum 200 kHz thanks to which it can coexist either in the Global System for Mobile Communications (GSM) spectrum or by occupying one of the legacy LTE Physical Resource Blocks (PRBs) as in-band or as guard-band. Since it coexists in the LTE spectrum, NB-IoT follows the legacy LTE numerologies as it uses OFDM) and SC-FDMA) in the downlink and uplink transmission schemes, respectively. Some modifications in the physical (PHY) and medium access control (MAC) layers are implemented to support the long-range massive machine-type (mMTC) connections with low power, low data rates,

low complexity, and hence low cost. However, despite its low complexity, this new radio access technology (RAT) delivers better performance in terms of the supported number of devices, and coverage enhancements for latency-insensitive applications with maximum coupling loss (MCL) of about 20 dB higher than LTE (i.e., 164 dB) [5, 6].

Let's see number of massive of device support in a cell in consideration.

As per the IoT requirements, there will be huge number of connected devices supporting different applications. NB-LTE needs to support this massive IoT capacity by using only one PRB in both uplink and downlink. NB-LTE with one PRB supports more than 52500 UEs per cell.

Inter-site distance (ISD) = 1732m

Cell site sector radius, *R = ISD/3 = 577.3m*

Acs- area of cell site sector (assuming a regular hexagon)

*Acs =3*sqrt(3/2)*R^2 = 0.866 sq km*

Hd -household density per sq km , Hd =1517

Nh-number of devices within a household, Nh =40

Ncs -number of devices per cell site sector

*Ncs = Acs * Hd * Nh = 0.866*1517*40= 52549 user/cell site.*

There is no establishment cause for delay tolerant traffic, because in NB-LTE all traffic is assumed to be delay tolerant. The high concentration of users in the cell radius requires an increase in cell bandwidth in conditions of high concentration of users of intelligent services NB-LTE network.

3GPP combines a variety of mobile technologies - from corporate picocell structures housed inside buildings to global satellite coverage. Thus, we can talk about the hierarchical structure of 3G networks. Pico and microcells are designed to serve slow-moving subscribers, while macrocells and satellite coverage areas are designed to serve subscribers at high and very high speeds. Microcells and macrocells mean the structural elements of neighboring levels in the hierarchy of the LTE mobile network, microcells are macrocells with respect to the picocell, which is a LTE microcell with respect to it. Subscribers divided by types of speed: fast, served by macrocells, and slow, served by microcells. For the design of hierarchical mobile networks, several templates use is the division of the network into clusters of cells and covering each cluster with one macrocell. In addition to serving NB IoT LTE users with different speed characteristics, the macrocell can also act as an additional resource in relation to the microcells. This means that if the microcell lacks channels to service incoming calls (new or transferred from anather microcells in the handover process), it can transfer the incoming call for service to the macrocell. This technology allows to quickly be responding to changes in subscriber load in the coverage area of the macrocell.

When using a macrocell as a resource that can shared, the access policy that most closely matches the dynamics of the change in the load on the microcells should apply. The macro cell can also be a dedicated resource, ie it may not serve fast moving subscribers, but only provide its own channels to service calls blocked in microcells. It is clear that the most efficient use of macrocell capacity to service subscriber calls is provided by using the capacity of the microcells themselves, when the free capacity of the macrocell remains maximum. To do this, the method of repackaging channels used, i.e. channels occupied by calls from subscribers of each microcell in the macrocell are pulled back into the microcell when one of its channels is released.

For evaluation of the change in the throughput of NB IoT LTE depending traffic intensity of macro- and microcells in NB-IoT LTE network proposed following mathematical model.

Consider a cluster of a two-tier LTE network consisting of *M* microcells and one macrocell (fig.2). The macrocell covers all M microcells. Microcell k has c_k communication channels for servicing subscriber calls, $k = \overline{1, M}$.

Subscriber calls occur in microcells *k* with intensity λ_k. and disappear (either at the end of the call, or when transferring a call to one of the neighboring cells) with intensity, $\mu_i \theta_{ji}$.

The end of the conversation in the k -th microcell occurs with probability P. The subscriber who leaves the service area of $\mu_i \theta_{ji}$. the microcell k, with probability moves to the service area of $\mu_i \theta_{ji}$the microcell $j, j = \overline{2, M}$. A macrocell has C communication channels that are used by

microcells to service incoming calls if all of the microcell's own channels are busy,is using C channels by macrocells of a two-tier LTE network cluster. The network of the k-th microcell is bounded by the threshold r_k, $k = \overline{1,M}$ A call that cannot be received due to the lack of channels in the corresponding microcell and in the macrocell is lost, and its impact on the call flow is not taken into account. When one of the channels is released into the microcells, the channels are repackaged with some intensity. To describe in the framework of the proposed model of calls coming from fast-moving subscribers directly to the macro cell, can put:

$$c_1 = 0, r_1 = C, \theta_{1j} = 0, \theta_{j1} = 0,$$

$j = \overline{2,M}$.

Figure 3 shows macro-and microcells NB-IoT LTE network of traffic intensity model.

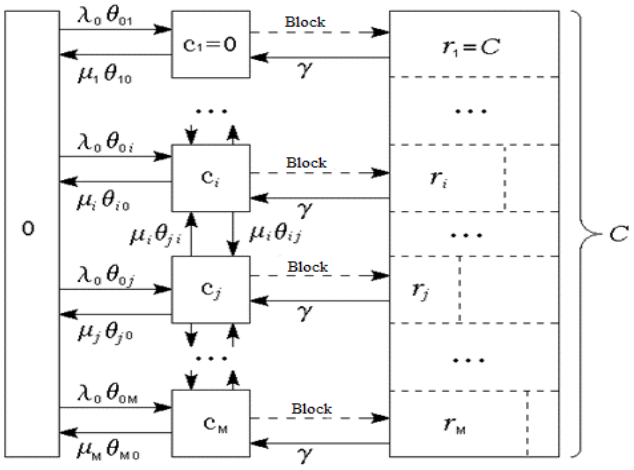


Fig. 3. A model describing of macro-and microcells NB-IoT LTE network traffic intensity

In this case determine $\lambda_0\theta_{01}$ and $\mu_1\theta_{10}$, respectively, the intensity of access to the macro cell and the intensity of service calls from fast-moving subscribers. Note that such a change means a decrease per unit number of microcells covered by the macrocell. The accepted model assumes that the average duration of a conversation in different microcells is different.

This allows can to flexibly describe the operating conditions of the simulated NB-IoTLTE network. For the case of instantaneous repackaging of channels ($\gamma < \infty$), the state space of the Markov process, which describes the operation of the system, has the form:

$$S := \left\{ n = (n_1, ..., n_m) : 0 \le n \le c + r, \sum_{k=1}^{M} (n_k - c_k)^+ \le C \right\}, \quad (1)$$

where $(x)^+ = \begin{cases} x, x \ge 0, \\ 0, x < 0. \end{cases}$

Can introduce the indicator function

$$x(n) = \begin{cases} 1, n \in S \\ 0, n \notin S \end{cases}$$

and $e_k := (I)_k$ – the k -th column of a unit matrix of dimension M.

Let's mark $\overrightarrow{\lambda_o} := (\lambda_0\theta_{01}, \ldots, \lambda_0\theta_{0M})$, let $\overrightarrow{\lambda} = (\lambda_1, \ldots, \lambda_M)$- solution of the equation, $\overrightarrow{\lambda}^T (E - \Theta) = \overrightarrow{\lambda_0^T}$, where $[\Theta]_{ij} := \theta_{ij}, i, j = \overline{1,M}$.

With some additional limitations, the described system has a convenient for analysis multiplicative solution [7]:

$$p(n) = G \prod_{k=1}^{M} \frac{p_k^{n_k}}{n_k!}, \; p_k = \frac{\lambda_k}{\mu_k}, \; k = \overline{1,M}, \; G = \left(\sum_{n \in S} \prod_{k=1}^{M} \frac{p_k^{n_k}}{n_k!} \right)^{-1}. \quad (2)$$

For $y = \infty$, well-known algorithms based on recursion or convolution can be used [8]. Consider the case of non-instantaneous repackaging of channels, when $y < \infty$.

Can introduce a random variable $X_{k1}(t), t \geq 0$ - the number of calls served by the microcell k at time t, and a random variable $X_{k1}(t), t \geq \theta$ - number of calls related to the microcell k, which at time - t is served by the macrocell, $k = \overline{1, M}$.

The operation of the system in this case can be described by a Markov process of dimension 2M. The process state space X (t) has the form:

$$S := \left\{ n = (n_{10},...,n_{M0}, n_{11},...,n_{M1} \right\} : n_{k0} \leq c_k,\ n_{k1} \leq r_k,\ k = \overline{1, M}\ ,\ \sum_{k=1}^{M} n_{k1} \leq C \quad (3)$$

.

To analyze the proposed model $y < \infty$, the theory of overload and the method of equivalent substitutions can used [9, 10].

The numerical analysis of the presented model for $y = \infty$ and $y = 0$ is performed below.

For considered a cluster of a two-tier mobile network (fig. 2): microcells - regular hexagons inscribed in a circle with a radius of 425 m; the radius of the macrocell is 1275 m.

The intensity of calls in each microcell per unit area is a in the macrocell: $1{,}5 \cdot 10^{-6}\ 1/s \cdot м^2\ \left(\lambda_{0i} = 0{,}8541/s, i = \overline{2{,}8}\right)$, of fast-moving subscribers - $10^{-7}\ 1/s \cdot м^2\ \left(\lambda_{0i} = 0{,}5251/s, i = \overline{2{,}8}\right)$, average talk time – 100s.

Each microcell has 8 channels for handling incoming calls. When using macrocell channels, there is a complete sharing policy, where $r_i = C, i = \overline{1{,}8}$.

According to the model there is a linear increase in throughput depending on the capacity of the macrocell C when using the general resource of the macrocell for a microcell of equal size without prior repacking of channels when servicing moving subscribers. Microcells have a uniform input load, so the intensity of occupancy of macrocell channels by calls belonging to different microcells is the same, i.e. an increase in the capacitance of a macrocell along the C-channels is equivalent to an increase in the capacitance of each microcell along the [C/M] channels. Channel repacking technology significantly increases system throughput. In the absence of moving subscribers and C=4, servicing of an additional load of 7 Earl is provided. That is, the efficiency of using a macrocell increases almost 3 times.

Cell selection and mobility procedure of macro and microcells NB-IoT LTE of cluster

NB-IoT LTE network is designed for infrequent and few byte data transmission between the UE and the network. It is assumed that the UE can exchange this information while being served from one cell, therefore, a handover procedure during RRC_CONNECTED is not needed. If such scenario a cell change would be required the UE has first go to the RRC IDLE state and re-select another cell there in.

For the RRC_IDLE state, cell re-selection is defined for both, intra frequency and inter frequency cells. Inter frequency refers here to the 180 KHz carrier, which means that even if two carriers are used in the in-band operation embedded into the same LTE carrier, this is still referred to as an inter-frequency re-selection. In order to find a suitable cell, the UE first measures the received power and quality of the NRS. These values are then compared to cell specific thresholds provided by the NB-SIB.

The S-criteria states that if both values are above these thresholds, the UE considers itself to be in coverage of that cell. If the UE is in coverage of one cell, it camps on it. Depending on the received NRS power, the UE may have to start a cell re-selection. The UE compares this power to a re-selection threshold, which may be different for the intra-frequency and the inter-frequency case. All required parameters are received from the actual serving cell, there is no need to read NB-SIBs from neighbors' cells. If multiple cell fulfill the S-criteria, the UE ranks the cells with respect to the

power excess over another threshold. A hysteresis is added in order to prevent too frequent cell re-selection.

Unlike conventional LTE, there are no priorities for the different frequencies. The UE finally selects the highest ranked cell, which is suitable, i.e. from which it may receive normal service. When the UE leaves RRC_CONNECTED, it does not necessarily select the same carrier to find a cell to camp on. The RRC Connection Release message may indicate the frequency on which the UE first tries to find a suitable cell. Only if the UE does not find a suitable cell on this frequency, it may also try to find one on different frequencies.

Conclusions

Has been analysis of Narrowband Internet of Things (NB-IoT) which allows enables a wide range of new IoT devices and services connected to the cellular network. Is shown then, NB-IoT is designed for fixed devices with low data transmission, low consumption, which leads to an increase in the number of devices connecting to each other. In turn, the massive NB-IoT modules that attempt to simultaneously request radio channel resources for uplink data transmission may suffer from random access preamble collision.

An increase in the efficiency of using the bandwidth of networks based on macro- and microcells with a high concentration of users of NB-IoT LTE networks is shown.

The results of a numerical analysis are presented to identify the factors affecting the system performance.

There is a linear increase in throughput depending on the capacity of the macrocell C when using the general resource of the macrocell for a microcell of equal size without prior repacking of channels when servicing moving subscribers, channel repacking significantly increases system throughput. In the absence of moving subscribers servicing of an additional load is provided, the efficiency of using a macrocell increases almost 3 times.

NB-IoT development will continue and will be expanded to include positioning methods, multicast services such as software updates or group-wide messages, mobility and service continuity, and additional technical details to expand the scope of NB-IoT technology.

References:

1 3GPP TR 45.820 Cellular system support for ultra-low complexity and low throughput Internet of Things (IoT).

2. 1/3GPP, TS 38.211, NR; physical channels and modulation, R16, v16.1.0, 2020.

3. Qi Pan, Xiangming Wen, Zhaoming Lui. Cluster-based Group for Massive Machine Type Communications under 5G Networks: DOI 10.1109/ACCESS.2020.2878424, IEEE Access, 2020, pp. 1–14.

4. RAN approved REL-13 NB_IoT CRs (RAN#72) Machina Research, May 2015.

5. Mekki K.; Bajic E.; Chaxel F.; Meyer F. A comparative study of LPWAN technologies for large-scale IoT deployment // ICT Express 2019. №5. P. 1–7.

6. Sinha R.S., Wei Y., Hwang S.H. A survey on LPWA technology: LoRa and NB-IoT // ICT Express 2017, 3, 14–21. Sensors 2019, 19, 2613 30 of 34.

7. 3GPP Specification: 36.932. Scenarios and Requirements for Small Cell Enhancements for E-UTRA and EUTRAN // Access :: http://www.3gpp.org/dynareport/ 36932.htm, 2013.

8. Berlin A. Digital cellular communication systems. 2007.

9. Horn G. 3GPP Femtocells: Architecture and Protocols. 2019. Access: https://www.qualcomm.com/documents/3gpp- femtocells-architecture-and-protocols.

10. Load Balancing function between two WANs ports on ZyWALL. Access: http://zyxel.ru/kb/1443.

Надійшла до редколегії 03.06.2024

Відомості про авторів:

Кадацька Ольга Йосипівна – канд. техн. наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри інфокомунікаційної інженерії ім. В.В. Поповського; Україна; e-mail: olga.kadatska@nure.ua; ORCID: https://orcid.org/0000-0002-5331-4324

Сабурова Світлана Олександрівна – канд. техн. наук, Харківський національний університет радіоелектроніки, доцент кафедри інфокомунікаційної інженерії ім. В.В. Поповського; Україна; e-mail: svitlana.saburova@nure.ua; ORCID: https://orcid.org/0000-0003-2214-2440

УДК 621.396.004 **DOI:10.30837/rt.2024.2.217.10**

Л.О. ТОКАР, канд. техн. наук, В.Є. ЦИЛЮРИК, В.В. СОЛОДІЛОВ

ДОСЛІДЖЕННЯ ПРОЦЕСУ РЕПЛІКАЦІЇ ДАНИХ ЗА ДОПОМОГОЮ АЛГОРИТМА РЕПЛІКАЦІЇ RAFT ДЛЯ ПІДТРИМКИ УЗГОДЖЕНОСТІ В КЛАСТЕРІ СЕРВЕРІВ

Вступ

Віртуальні IP-АТС є популярним рішенням для організацій, які хочуть мати гнучкість та контроль своїх комунікаційних сервісів. Для забезпечення максимально ефективного та надійного використання технологій віртуальних IP-АТС у складних розподілених середовищах, а також для підвищення якості послуг та зниження загроз зупинки роботи системи у випадках збоїв або атак, використовуються методи та механізми, які є вирішальними для забезпечення стійкості комунікаційних сервісів.

Методи та технології забезпечення високої доступності та відмовостійкості для віртуальних сервісів є важливими компонентами сучасних інформаційних структур. Завдяки їм компанії можуть забезпечувати безперебійну роботу своїх систем і служб, незалежно від можливих відмов та збоїв [1]. Вирішення питань забезпечення стійкої роботи віртуальних IP-АТС полягають у кластерізації серверів, використанні методів контейнеризації, механізмів управління та контролю та механізмів реплікації для забезпечення консистентності даних у кластері.

Для забезпечення відмовостійкості використовуються резервне копіювання та реплікація даних. Це дозволяє відновлювати систему після відмови, не втрачаючи важливої інформації. Висока доступність забезпечується за допомогою реплікації даних, автоматичного перемикання між серверами або дата-центрами та швидкого відновлення після відмов. Реплікація не тільки підвищує надійність системи, але й покращує продуктивність, оскільки запити до даних можуть оброблятися локально на кожному вузлі, зменшуючи час відгуку та навантаження на мережу.

Враховуючи важливість реплікації даних, особливу увагу слід приділити безпеці. Репліковані дані повинні бути захищені від несанкціонованого доступу та змін, особливо у випадках, коли вузли кластера розташовані в різних фізичних місцях або використовуються хмарні послуги. Це вимагає впровадження сильних механізмів автентифікації, шифрування та контролю доступу. З огляду на сучасні вимоги до масштабованості та доступності великих розподілених систем, реплікація є ключовим компонентом, що забезпечує ефективну та надійну роботу додатків та послуг. Це й обумовлює актуальність даної роботи.

Кластеризація, яка дозволяє групувати кілька серверів у єдиний віртуальний ресурс, є одним з ключових методів досягнення високої доступності. Кластеризація на основі реплікації є ключовим елементом у сучасних розподілених системах, забезпечуючи високий рівень доступності та відмовостійкості. Основна ідея полягає у створенні декількох копій даних на різних вузлах кластера, що дозволяє системі продовжувати працювати без перебоїв навіть у випадку збоїв або відмов окремих компонентів [2].

Окрім вибору стратегії реплікації, важливим аспектом є управління навантаженням та моніторинг стану різних вузлів у кластері. Ефективне розподілення навантаження між вузлами може значно покращити загальну продуктивність та доступність системи. Це включає не тільки рівномірний розподіл запитів між вузлами, але й адаптацію до змін у навантаженні та можливих відмовах компонентів. Моніторинг стану вузлів дозволяє оперативно виявляти та реагувати на проблеми, забезпечуючи безперервну роботу системи.

Сучасні рішення для забезпечення високої доступності та відмовостійкості також включають в себе георедунтантність, коли дата-центри розташовані в різних географічних локаціях. Це не тільки розподіляє ризики, але і дозволяє забезпечити безперебійний доступ

до ресурсів для користувачів з різних регіонів. Прикладом є розгортання дата-центрів в Amazon Web Services (AWS).

Важливо зазначити роль програмного забезпечення у забезпеченні високої доступності та відмовостійкості. Сучасні системи управління даними, такі як бази даних та системи управління контентом, часто включають в себе механізми реплікації, кешування та резервного копіювання, що допомагає збільшити стабільність та доступність даних [3].

Забезпечення високої доступності та відмовостійкості потребує комплексного підходу, який поєднує апаратні, програмні та організаційні рішення. Важливими моментами є регулярність в оновлювленні технологій, проведення аудиту систем на предмет потенційних слабких місць й навчання персоналу найкращим практикам у галузі забезпечення безперебійної роботи віртуальних інфраструктур.

Основна частина

Реплікація даних є важливою стратегією для забезпечення надійності, доступності та продуктивності систем обробки даних. Цей процес полягає у створенні та підтримці копій даних на різних серверах або вузлах, щоб забезпечити можливість доступу до даних у випадку відмови обладнання та програмного забезпечення чи інших непередбачуваних подій.

Реплікація може бути використана для забезпечення доступу до даних у випадку відмови сервера, підвищення продуктивності через розподілення навантаження або для забезпечення географічної резервності даних.

Відомі різні типи реплікації: повна реплікація, часткова реплікація, синхронна реплікація, асинхронна реплікація, які відрізняються способами копіювання та різним ступенем доступності даних.

Високий рівень відмовостійкості і доступності даних забезпечує метод реплікації на рівні блоків, що дозволяє створити точні копії даних на різних пристроях. У цьому методі зміни, що відбуваються у блоку даних на одному диску, автоматично реплікуються на інший диск. У хмарних сервісах дані також можуть реплікуватися на різні сервери [4].

Даний метод широко використовується у системах Redundant Array of Independent Disks (RAID), які гарантують, що навіть у разі відмови одного диску дані залишаються доступними та незмінними (рис. 1).

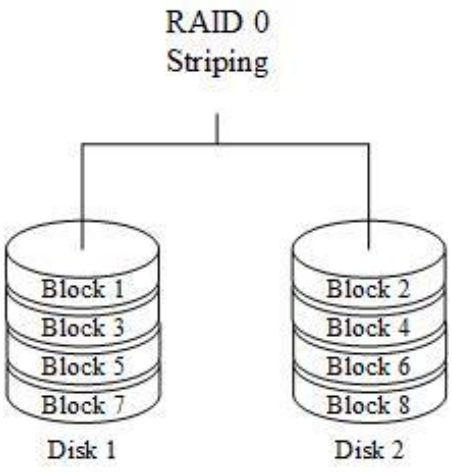


Рис. 1. Діаграма блочної реплікації у системі RAID

Такий підхід до реплікації має широкий спектр застосувань. Його може бути використано для оптимізації роботи з базами даних чи в системах з розподіленою файловою системою, де реплікація блоків даних допомагає забезпечити надійність та швидкодіючий доступ до інформації.

Об'єктна реплікація – це метод реплікації даних, який копіює або синхронізує об'єкти даних (які можуть включати файли, блоки даних чи інші структуровані формати) між різними системами зберігання (рис. 2).

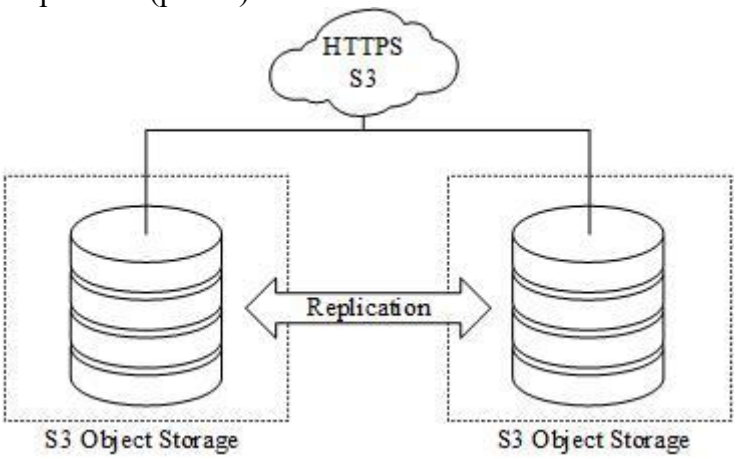


Рис. 2. Діаграма об'єктної реплікації

Цей метод реплікації широко використовується в хмарних та розподілених системах зберігання, де необхідно забезпечити надійність та доступність великих обсягів даних. Він дозволяє легко масштабувати системи зберігання й забезпечує швидкий доступ до даних, але також може вимагати складнішого управління та забезпечення додаткового простору для зберігання.

Кожний з методів реплікації має свої переваги та обмеження, тому є певний компроміс між стійкістю отриманих даних та проблемами обмеження в швидкості чи в масштабуванні. Цілісність даних, отриманих в процесі реплікації, може обмежити швидкість запису. Легкість в масштабуванні системи зберігання й швидкий доступ до даних може вимагати складнішого управління та забезпечення додаткового простору для зберігання. Зменшення ризику втрати даних і забезпечення безперервної роботи послуг для користувачів вносять проблеми керування конфліктами, які виникають при оновленні реплік і потребують розробки механізмів для їх вирішення. Тому важливо вибрати підхід, що найкраще відповідає специфічним вимогам та характеристикам системи.

Розподілені системи, як важлива частина сучасних обчислювальних мереж, вимагають надійних механізмів для досягнення консенсусу та управління даними. Алгоритм Raft, розроблений як альтернатива складнішим алгоритмам, таким як Paxos, забезпечує ефективне та зрозуміле рішення для реплікації журналу в розподілених системах. Цей алгоритм вирішує проблему досягнення консенсусу, забезпечуючи такі умови, що кожний вузол в системі може безпечно та надійно виконувати однакові операції у встановленому порядку.

В роботі проведено аналіз літератури з використання алгоритму консенсусу Raft. В роботі [55] запропоновано використання алгоритму консенсусу Raft як більш зрозумілого й модульного підходу в порівнянні з іншими алгоритмами консенсусу.

Ймовірність поділу розподіленої мережі з використанням простого, але точного аналітичного підходу, проаналізовано у роботі [6]. Ймовірність поділу мережі представлено як функцію розміру мережі, швидкості втрати пакетів та тривалості тайм-ауту. Для перевірки використано симулятор алгоритму Raft. За допомогою розробленої моделі запропоновано теоретично передбачити час та ймовірність поділу мережі, а також оптимізувати параметри алгоритму консенсусу Raft.

Модель для імітації системи Blockchain з використанням теорії масового обслуговування запропоновано у роботі [7]. Модель створено з використанням СМО M/M/1 у якості пулу пам'яті та СМО M/M/c у якості пулу майнінгу. Цей метод простий, але ефективний для виявлення важливих показників мережі: при визначенні кількості транзакцій на блок, часу майнінгу кожного блоку, пропускної здатності системи/транзакцій в секунду, кількості пулів

пам'яті, часу очікування в пулі пам'яті, кількості непідтверджених транзакцій у всій системі, загальної кількості транзакцій та кількості згенерованих блоків.

Модифікований консенсусний алгоритм стійкості до відмови для вирішення проблем масштабованості Istanbul Byzantine Fault Tolerant (IBFT) в приватних Blockchain-системах розглянуто в роботі [8]. Використано інструмент для тестування Blockchain Hyperledger Caliper, який дозволяє оцінити продуктивність платформ з відкритим вихідним кодом, таких як Quorum та Hyperledger Fabric. Використання Fabric протоколів Solo і Raft як служби замовлення транзакцій у версії 1.4 забезпечує надійну пропускну здатність та низьку затримку, що робить Hyperledger Fabric хорошим кандидатом для корпоративного рішення на основі Blockchain. З іншого боку, реалізація Raft та IBFT у Quorum як протоколів консенсусу, які відповідають за підтримку ланцюжка блоків, робить його більш логічним варіантом для тестування модифікацій.

В роботі [9] запропоновано механізм вибору лідерів за допомогою алгоритму Raft з урахуванням стабільності мережі. Розроблений метод знизив можливість поділу мережі за допомогою Raft. Крім того, це дало змогу запобігти подальшому зниженню продуктивності, коли мережа вузлів Blockchain нестабільна.

У роботі [10] розглянуто функціональні переваги та недоліки алгоритму Raft; характеристики безпеки, довіри між учасниками, пропускної здатності та масштабованості. Виявлено, що жоден алгоритм, окрім Raft, не продемонстрував повного домінування у всіх аспектах порівняння.

У роботі [11] розглянуто СМО M/G/1 з ланцюгом Маркова за дискретним часом Discrete-Time Markov Chain (DTMC) у системі Blockchain. За допомогою теорії масового обслуговування теоретично проаналізовано процеси генерації блоків та побудови системи Blockchain. За допомогою мови програмування Python проведено знаходження середнього часу підтвердження транзакції, кількості транзакцій у черзі у точці прибуття, кількості транзакцій у точках відправлення та інші параметри.

Крім того, розглянуто марківську модель СМО Mb/M/1 як теоретичну основу для аналізу системи Blockchain на основі доказу повноважень, що фокусуються на стохастичній поведінці транзакцій та для аналізу консенсусного протоколу повноважень. Представлено дані, які проведені на Python. Зазначено, що оптимальна пропускна здатність блоку досягається при невеликій кількості транзакцій.

У контексті розподілених систем, реплікація даних є ключовим фактором для забезпечення високої доступності та стійкості системи. Raft дозволяє системам ефективно керувати реплікацією, забезпечуючи консистентність даних між вузлами. Це особливо важливо в сценаріях, де втрата або неконсистентність даних може призвести до серйозних наслідків.

Алгоритм Raft базується на концепції лідера. Всі вузли в кластері є або послідовниками (followers), або кандидатами, поки один з них не стає лідером. Лідер керує процесом реплікації даних, відправляючи оновлення журналу іншим вузлам. Такий підхід спрощує процес реплікації та знижує ймовірність розбіжностей даних.

Розробка алгоритму Raft вмотивовано потребою в більш зрозумілому та менш складному рішенні для досягнення консенсусу в розподілених системах, ніж наявні алгоритми. З часом Raft здобув популярність у спільноті розробників завдяки своїй простоті та ефективності. Його чітка структура та легкість у розумінні зробили його популярним вибором для багатьох великих розподілених систем.

Зазвичай архітектура розподілених систем базується на принципах мікросервісів та контейнеризації. Механізмом управління та контролю є платформа Kubernetes. Kubernetes відповідає за управління цими контейнерами, гарантує їх доступність та відмовостійкість.

У контексті Kubernetes, алгоритм Raft може бути використано для керування розподіленою системою еталонів (etcd), яка є ключовою складовою для зберігання та синхронізації конфігурації кластера Kubernetes. Etcd використовує Raft для забезпечення

сильної консистентності (strong consistency) стану кластера, що є критично важливим для забезпечення надійної та передбачуваної оркестрації контейнерів [12].

Raft дозволяє кластеру etcd працювати у якості єдиного цілісного компоненту, незважаючи на фізичну розподіленість між вузлами [13]. Це досягається завдяки реплікації журналу команд між всіма вузлами, що входять до кластера etcd, забезпечуючи сильну консистентність та високу доступність. Консенсусну архітектуру etcd з використанням алгоритму Raft показано на рис. 3.

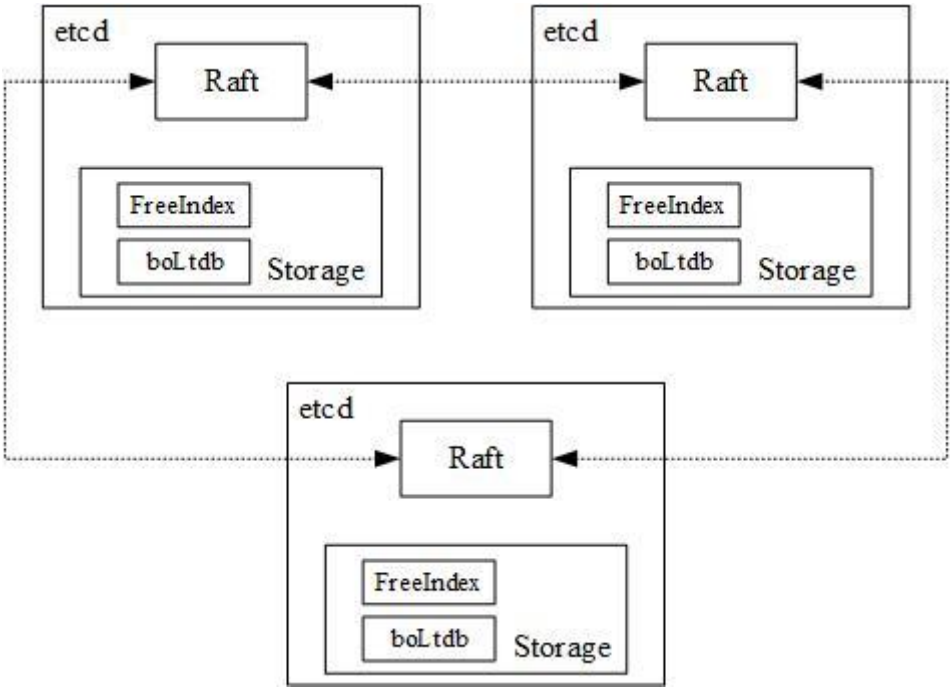


Рис. 3. Консенсусна архітектура etcd з використанням алгоритму Raft

Надійність Kubernetes в значній мірі залежить від надійності etcd, оскільки etcd зберігає важливий стан, що включає конфігурацію мережі, сервіси, політики безпеки та інше. Якщо дані в etcd не будуть консистентними або якщо сервіс стане недоступним, то це може призвести до серйозних збоїв в роботі кластера.

Raft гарантує, що навіть у разі збоїв у кількох вузлах, etcd зможе зберігати стан кластеру без втрат. Коли лідер отримує зміни стану, він не застосовує ці зміни до свого локального стану до тих пір, поки не отримає підтвердження, що зміни були репліковані на більшість вузлів. Такий підхід запобігає режиму split-brain, коли різні частини кластера можуть мати суперечливі стани.

Завдяки використанню etcd та Raft, платформа Kubernetes може забезпечити високий рівень стійкості та надійності, які необхідні для сучасних розподілених систем. Raft особливо корисний для середовищ, де потрібна сильна консистентність та здатність швидко відновлювати систему після збоїв.

Однією з основних характеристик Raft є його структура з чітко визначеними ролями (лідер, кандидат, послідовник) та простими правилами переходу між цими станами. Це робить процес лідерства та реплікації даних прозорим та передбачуваним. Крім того, використання випадкових тайм-аутів для ініціації виборів знижує ймовірність конфліктів та забезпечує ефективний вибір лідера.

На сьогодні Raft є одним з основних алгоритмів для досягнення консенсусу у розподілених системах. Його здатність забезпечувати надійну реплікацію даних та високий рівень доступності робить його незамінним інструментом в арсеналі розробників, що працюють із складними розподіленими системами.

Простота методу консенсусу алгоритму Raft робить його легшим для розуміння та застосування в порівнянні з іншими алгоритмами консенсусу. Також, Raft забезпечує кращі гарантії безпеки, ніж існуючі алгоритми консенсусу. Він гарантує, що система залишається послідовною, навіть коли відбувається кілька збоїв або розділів мережі. Відомі блокчейн-платформи, такі як R3 Corda та Quorum, що використовують Raft як протокол консенсусу.

У Raft лідер відіграє ключову роль у координації дій між різними вузлами кластера. Лідер не тільки відповідає за реплікацію журналу записів до інших вузлів, але й гарантує їх консистентність та порядок. Вибір лідера є критичним кроком у роботі Raft, оскільки від цього залежить стабільність та ефективність усієї системи.

Процес вибору лідера в Raft ініціюється, коли вузол переходить у стан кандидата. Це відбувається у випадку, якщо вузол не отримує достатньої кількості heartbeats від поточного лідера протягом певного тайм-ауту. Кандидат ініціює вибори, генеруючі голосування серед інших вузлів кластера.

Під час виборчого процесу, кандидат відправляє запити на голосування до інших вузлів у кластері. Щоб вузол став лідером, він повинен отримати більшість голосів від інших вузлів. У випадку, якщо декілька вузлів стають кандидатами одночасно, система може не досягнути консенсусу, що призведе до повторення виборчого процесу після наступного тайм-ауту. Діаграму станів вибору лідера в алгоритмі Raft показано на рис. 4.

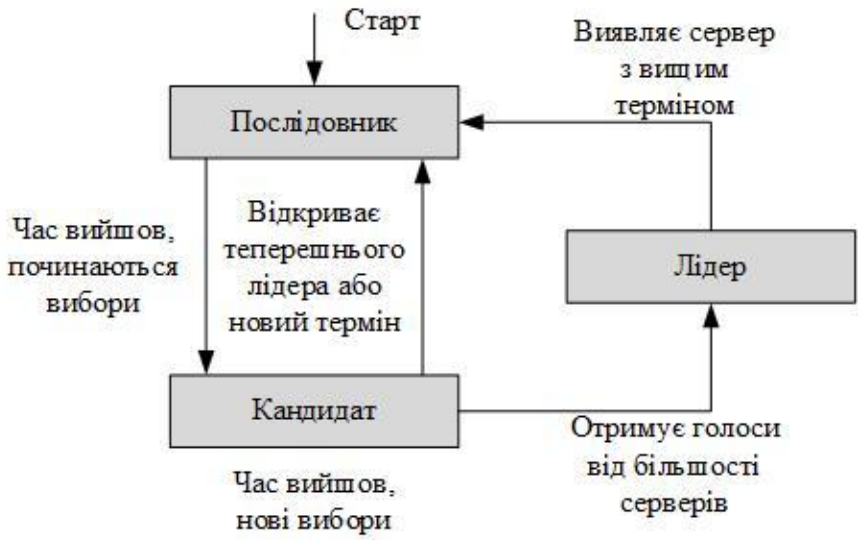


Рис. 4. Діаграма станів вибору лідера в алгоритмі Raft

Консистентність даних між вузлами в розподіленій системі в алгоритмі Raft забезпечує реплікація журналу, що є фундаментальним процесом. Цей процес відбувається під керівництвом лідера, який відповідає за зміни при веденні й синхронізацію журналу на всіх вузлах кластера. У журналі міститься серія записів, які вузли повинні включатися у визначеному порядку.

Після обрання лідера починається обслуговування запитів клієнтів. Кожен запит містить команду, яка повинна бути виконана на розподіленому консистентному автоматі. Лідер додає команду до свого журналу як новий запис, а потім паралельно для кожного сервера викликає процедуру AppendEntries, щоб реплікувати цей запис.

Лідер вирішує, коли можна виконати наступний запис у журналі розподіленого консистентного автомату. Ці записи називаються зафіксованими. Raft гарантує, що зафіксовані записи не будуть втрачено та в кінцевому підсумку будуть виконано на всіх доступних консистентних автоматах.

Raft гарантує наступні властивості: якщо два записи у різних журналах мають однаковий індекс і номер епохи, то вони містять одну й ту ж команду; якщо два записи у різних журна-

лах мають однаковий індекс і номер епохи, то всі попередні команди в цих журналах ідентичні. Невідповідності можуть накопичуватися в результаті серії відмов лідерів й вузлів.

Вузол може не мати записів, які є у лідера (так як лідер не встиг реплікувати всі записи), він може мати додаткові записи, яких немає у лідера (так як старий лідер почав реплікацію запису, але відмовив під час виконання цієї операції), а також можуть бути записи, яких немає ні в одному з серверів. У журналі вузла може бути кілька таких записів.

Таким чином, значення процедури AppendEntries досягне точки, де лідер й вузол мають ідентичні журнали. Коли це станеться, перевірка узгодженості AppendEntries успішно виконається, потім процедура видалить конфліктуючі записи в журналі вузла, а потім додасть записи з журналу лідера (якщо такі існують). Після успішного виконання процедури AppendEntries журнал вузла буде узгоджений з лідером.

В контексті роботи алгоритму Raft тайм-аути та терміни є двома фундаментальними концепціями, які сприяють стабільності та надійності реплікації даних в розподілених системах. Тайм-аути використовуються для ініціювання виборів та запобігання конфліктів, а терміни допомагають забезпечити послідовність й визначити часові рамки діяльності кластера.

Тайм-аути в Raft мають критичне значення в процесі виборів лідера. Коли вузол не отримує повідомлень від лідера протягом певного інтервалу часу (тайм-аут), він переходить у стан кандидата та ініціює нові вибори. Цей механізм гарантує, що кластер може швидко реагувати на втрату лідера та забезпечити безперервність управління даними.

Налаштування тайм-аутів є ключовим механізмом для забезпечення стабільності роботи кластера. Занадто короткі тайм-аути можуть спричинити часті та непотрібні вибори лідера, тоді як занадто довгі тайм-аути можуть затримувати відновлення системи після втрати лідера.

Процес консенсусу Raft використовується для підтримки узгодженості в кластері. Таким чином, Raft розкладає проблему консенсусу на відносно незалежні підпроблеми, які можна описати наступним чином [14]:

– вибори лідера: нового лідера потрібно обирати, коли існуючий лідер зазнає невдачі;

– реплікація журналу: лідер повинен приймати записи журналу від клієнтів і реплікувати їх по кластеру, змушуючи інші журнали погоджуватися з власним.

В контексті моделі черг процес реплікації журналу алгоритма консенсусу Raft розглядається як система черги з декількома серверами. В роботі використано модель СМО M/M/c для дослідження систем консенсусу, де прибуття моделюються як процес Пуасона, а час обслуговування має експоненційний розподіл. Модель M/M/c є однією з базових моделей у теорії масового обслуговування та використовується для аналізу систем з кількістю обслуговуючих пристроїв (серверів) більше одного. Термін M/M/c описує систему, де M відноситься до Markovian (експоненційний розподіл часів між прибуттям клієнтів та обслуговуванням), а c позначає кількість серверів у системі [15].

У процесі роботи для консенсусу Raft сервер відноситься до вузла-послідовника у розподіленій системі, який бере участь у протоколі консенсусу. Кожен сервер підтримує копію журналу системи та спілкується з іншими серверами, щоб переконатися, що вони погоджуються з поточним станом журналу. Запити клієнтів надходять до кожного вузла-послідовника згідно з процесом Пуасона з інтенсивністю λ, і вимагають часу обслуговування $1/\mu$.

Процес прибуття: як тільки лідера обрано, то він починає обслуговувати запити клієнтів. Кожен запит клієнта містить команду, яка має бути виконана реплікованими становими машинами. Лідер додає команду до свого журналу як новий запис.

Таким чином, процес прибуття моделюється як процес Пуасона з інтенсивністю, яку визначено частотою повідомлень лідера, які відправляються послідовникам. Показник λ – це інтенсивність прибуття запитів клієнта від вузла-лідера до вузлів-послідовників.

Процес обслуговування наступний: лідер додає команду до свого журналу як новий запис, потім видає Remote Procedure Call (RPC) запити AppendEntries паралельно кожному з інших послідовників для реплікації запису. Коли запис було безпечно репліковано, лідер застосовує запис до своєї станової машини і повертає результат цього виконання клієнту.

Час обслуговування також моделюється як процес Пуасона з експоненційним розподілом та з інтенсивністю, що визначено часом, який потрібний послідовникам для реплікації записів журналу лідера. Показник μ – це швидкість обслуговування кожного послідовника, що представляє швидкість, з якою записи журналу реплікуються від послідовника до лідера.

В роботі розроблено код програми для проведення експериментів на мові Python, фрагменти якого надано на рис. 5 та 6.

```
import numpy as np
import matplotlib.pyplot as plt

# Параметри, які згадані в завантаженому зображенні
mu_values = [625, 650, 675, 700]  # різні значення сервісної швидкості м
c_values = [3, 5, 11]             # різні значення кількості каналів с
lambda_range = np.arange(150, 601) # діапазон значень л

# Функція для розрахунку очікуваного часу запиту в системі (latency)
def expected_latency(lmbda, mu, c):
    if lmbda < c * mu:
        return 1 / mu
    else:
        return c / (lmbda * (c * mu - lmbda))

# Функція для розрахунку ймовірності затримки повідомлення
def probability_of_delay(lmbda, mu, c):
    return lmbda / (c * mu)
```

Рис. 5. Фрагмент коду програми на Python

```
# Графіки для різних с та м
for c in c_values:
    for mu in mu_values:
        latency = [expected_latency(lmbda, mu, c) for lmbda in lambda_range]
        axes[1, 1].plot(lambda_range, latency, label=f'c={c}, м={mu}')
axes[1, 1].set_title('Latency for different м and c')
axes[1, 1].set_xlabel('Arrival Rate л')
axes[1, 1].set_ylabel('Latency')
axes[1, 1].legend()
axes[1, 1].grid(True)
plt.tight_layout()
plt.show()
```

Рис. 6. Фрагмент коду програми на Python

Прийняття запитів клієнтів до середньої інтенсивності λ відповідно до моделі Пуассона визначено за формулою

$$\lambda_n = \lambda. \tag{1}$$

Існує c паралельних ідентичних послідовників (серверів), кожен з яких працює згідно з експоненційним розподілом з середньою швидкістю μ. Для n клієнтських запитів, швидкість обслуговування сервера може бути отримана в наступних ситуаціях:

– якщо $n \langle c$, то всі клієнтські запити можуть бути обслужено одночасно. Черга буде відсутня. Кількість простоюючих серверів у цьому випадку становить $c - n$. Тоді $\mu n = n\mu$, $n = 0, 1, 2, ..., c$;

– якщо $n \geq c$, то всі сервери зайняті та максимальна кількість клієнтських запитів, що очікують в черзі, становитиме $n - c$. Тоді $\mu_n = c_\mu$ для $n = 0, 1, 2, ...$

Таким чином, швидкість обслуговування μ_n може бути визначено:

$$\mu_n = \begin{cases} n\mu, & n \leq c \\ c\mu, & c \leq n \end{cases}. \tag{2}$$

Цей аналіз дає розуміння того, як алгоритм Raft працює в різних контекстах і може бути застосований для оптимізації процесів проєктування систем.

Продуктивність Raft може бути підвищена шляхом зміни характеристик системи, таких як, кількість вузлів-послідовників, їх обслуговуюча здатніть, надходження запитів від клієнтів тощо. Результати проаналізовано між компромісами пропускної здатності, затримки та ймовірності затримки. Це може допомогти системним дизайнерам робити практичні вибори щодо розподілу ресурсів та налаштування системи.

Ймовірність затримки повідомлення визначено:

$$P(\lambda, \mu, c) = \frac{\lambda}{c\mu}. \tag{3}$$

Очікувану затримку визначено:

$$L(\lambda, \mu, c) = \begin{cases} \frac{1}{\mu}, & \\ \frac{c}{\lambda(c\mu - \lambda)}, & \lambda \langle c\mu. \end{cases} \tag{4}$$

У роботі проведено моделювання в середовищі Visual code на Python, щоб продемонструвати зміну основних параметрів для відповідного механізму консенсусу Raft моделі технології блокчейн у контексті системи черги.

Для різних значень μ при $\lambda \in [150, 600]$ досліджено такі параметри: кількість послідовників, очікуваний час запиту клієнта у системі, тобто затримка, ймовірність затримки повідомлення тощо. На рис. 7 показано графіки обробки запитів клієнтів для $\mu = 650$.

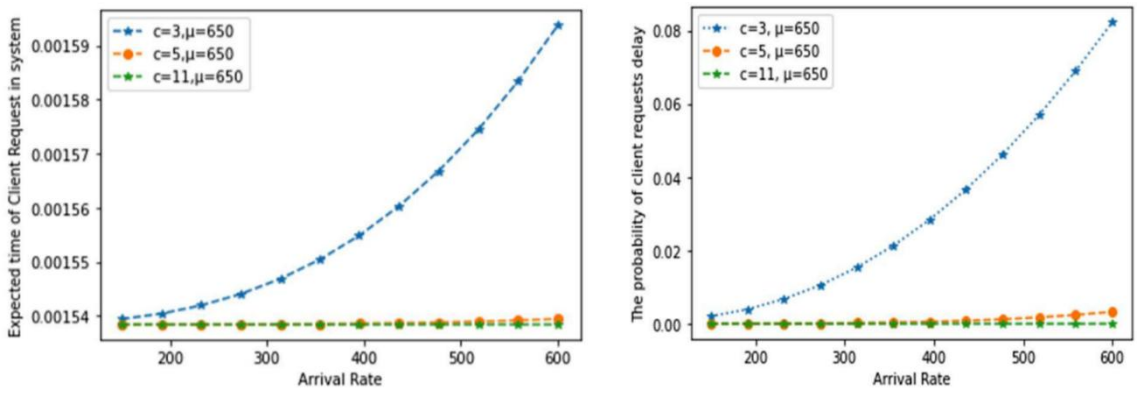


Рис. 7. Графіки обробки запитів клієнтів для $\mu = 650$

При аналізі графіків очікуваного часу запиту клієнта, який відноситься до затримки, та ймовірності затримки для $\mu = 650$, видно, що по мірі збільшення швидкості, очікуваний час запиту клієнта збільшується для $c = 3$.

Для $c = 5$ та $c = 11$ видно, що при більшій кількості послідовників параметр затримки та затримки повідомлень не залежать від швидкості прибуття.

На рис. 8 показано графіки обробки запитів клієнтів для інших значень μ.

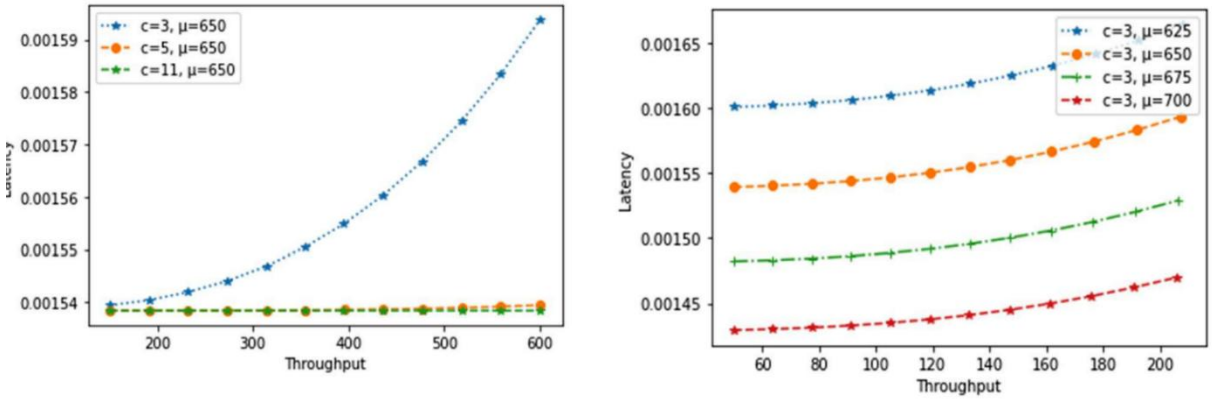


Рис. 8. Графіки очікуваного часу запиту клієнта для інших значень μ

При аналізі графіків очікуваного часу запиту клієнта, який визначається як затримка, та ймовірності затримки повідомлення для $c = 3$ та різних μ видно, що по мірі збільшення швидкості прибуття очікуваний час запиту клієнта зростає для $c = 3$ та різних μ.

На рис. 9 показано графіки аналізу очікуваного часу клієнтських запитів та ймовірності затримок у системі.

При аналізі графіків очікуваного часу запиту клієнта, який визначається як затримка, та ймовірності затримки повідомлення для комбінацій c та μ, видно, що по мірі збільшення швидкості прибуття очікуваний час запиту клієнта зростає для комбінацій c та μ.

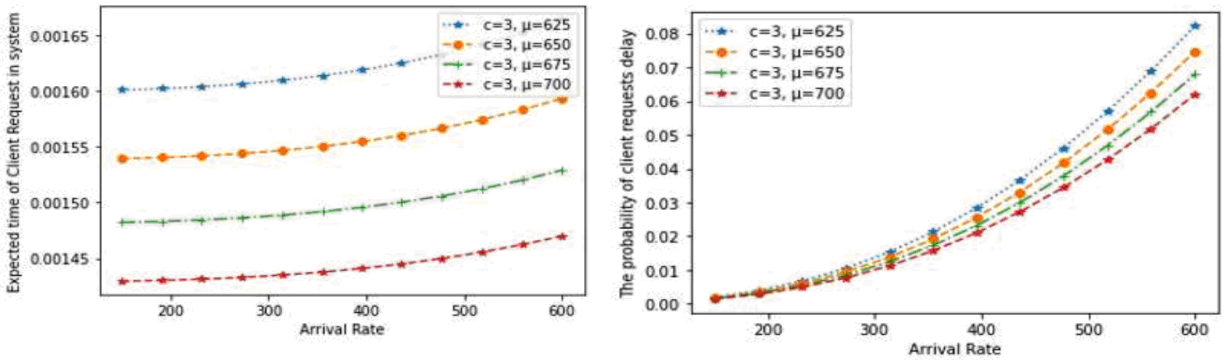


Рис. 9. Графіки аналізу очікуваного часу клієнтських запитів та ймовірності затримок у системі

Таким чином, проаналізовано показники продуктивності: очікуваний час запитів клієнтів у системі, затримка, очікувана кількість запитів клієнтів у системі та черзі, пропускна здатність, ймовірність затримки повідомлення алгоритму консенсусу Raft.

Висновки

В роботі наведено типи реплікації даних та обґрунтовано її використання для забезпечення надійності та доступності даних в розподілених системах. Проведено аналіз літератури

з використання алгоритму консенсусу Raft та наведено основи цього алгоритму. Проведено аналіз математичної моделі процессу консенсусу Raft, використано модель СМО M/M/c.

Проведено аналіз моделі M/M/c, яка є фундаментальною в теорії масового обслуговування, в якій час між прибуттям заявок та час обслуговування підпорядковуються експоненційному розподілу. Ця модель дозволяє оцінити ефективність системи кластерізації та реплікації Raft з кількома каналами обслуговування.

Проведено моделювання в середовищі Visual code на Python для дослідження зміни основних параметрів для відповідного механізму консенсусу Raft моделі технології блокчейн у контексті системи черги. Отримано графіки очікуваного часу запиту клієнта та ймовірність затримки у системі для різних значень швидкості обслуговування та кількості серверів. Доведено, що по мірі збільшення швидкості прибуття очікуваний час запиту клієнта зростає.

Список літератури:

1. DevOpsCube. How to Setup Prometheus Monitoring On Kubernetes Cluster. URL: https://devopscube.com/setupprometheus-monitoring-on-kubernetes.
2. Л.О. Токар, О.А. Колтаков, В.Є. Цилюрик. Створення тестового стенду Call-центру для балансування навантаження серверів Asterisk у кластері // Радіотехніка. 2023. № 212. С. 186–196. doi:10.30837/rt.2023.1.212.18.
3. Grafana Dashboards. Node Exporter Full. URL: https://grafana.com/grafana/dashboards/1860.
4. P. Pyda, M. Przywuski, T. Dalecki, J. Sliwa. Efficiency of Virtual Machine Replication in the Data Center // International Conference on Military Communications and Information Systems (ICMCIS 2022). 2022. Vol. 205. P. 208–217. doi:10.1016/j.procs.2022.09.022.
5. R.A. Memon, J.P. Li, J. Ahmed. Simulation Model for Blockchain Systems Using Queuing Theory // Electronics. 2019. Vol. 8 (2). P. 1–19. doi:10.1007/978-3-031-21229-1_1.
6. A. Baliga, I. Subhod, P. Kamat & S. Chatterjee. Performance evaluation of the quorum blockchain platform. URL: https://arxiv.org/abs/1809.03421.
7. Q.L. Li, J.Y. Ma, Y.X. Chang. Blockchain Queue Theory // Proceedings of the 7th International Conference on Computational Social Networks, CSoNet 2018. 18-20 Dec., 2018. Vol. 11280. P.25-40. doi:10.1007/978-3-030-04648-4_38.
8. D. Ongaro & J. Ousterhout. In search of an understandable consensus algorithm // 2014 USENIX Annual Technical Conference. 2014. P. 305–319. doi: 10.25209/2079-3316-2021-12-2-137-192.
9. D. Kim, I. Doh & K. Chae. Improved raft algorithm exploiting federated learning for private blockchain performance enhancement // 2021 IEEE International Conference on Information Networking (ICOIN). 2021. P. 828–832. doi: 10.1109/ICOIN50884.2021.9333932.
10. S. Vora, N. Thakkar, R. Gor. A Study of Performance Measures and Throughput of Raft Consensus Algorithm // International Journal for Research in pplied Science & Engineering Technology (IJRASET). 2023. Vol. 45. P. 862–869. doi:10.22214/ijraset.2023.54751.
11. G. Yang, K. Lee, Y. Yoo, H. Lee & C. Yoo. Resource Analysis of Blockchain Consensus Algorithms in Hyperledger Fabric // IEEE Access. 2022. Vol. 10. P. 74902–74920. doi: 10.1109/ACCESS.2022.3190979.
12. Q. Huo, S. Li, Y. Xie and Z. Li. Horizontal Pod Autoscaling based on Kubernetes with Fast Response and Slow Shrinkage // 2022 International Conference on Artificial Intelligence, Information Processing and Cloud Computing (AIIPCC), Kunming, China, 2022. P. 203–206. doi: 10.1109/AIIPCC57291.2022.00051.
13. A. Vaghani, K. Sood, S. Yu. Security and QoS issues in blockchain enabled next-generation smart logistic networks: A tutorial Blockchain // Research and Applications. 2022. Vol.3. P. 1–14. doi: 10.1016/j.bcra.2022.100082.
14. С. Журавель, О. Шпур, Ю. Пиріг. Досягнення консенсусу в розподілених сервісних системах // Інфокомунікаційні технології та електронна інженерія. 2022. № 2 (4). С. 58–66. doi:10/23939/ictee2022.02.058.
15. В.В. Гнатушенко, Г.К. Витовтов. Аналіз систем масового обслуговування при стрибкоподібній зміні інтенсивностей потоків інформації // Прикладні питання математичного моделювання. 2021. Т. 4, no 2.1. С. 76–83. doi: 10.32782/KNTU2618-0340/2021.4.2.1.7.

Надійшла до редколегії 02.06.2024

Відомості про авторів:

Токар Любов Олександрівна – канд. техн.. наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри інфокомунікаційної інженерії імені В.В. Поповського, Україна; email: liubov.tokar@nure.ua; ORCID: https://orcid.org/0000-0002-7780-1928

Цилюрик Вадим Євгенович - компанія IT-Lance, Україна, email: vadym.tsyliuryk@nure.ua

Солоділов Віктор Васильович – Харківський національний університет радіоелектроніки, магістр кафедри інфокомунікаційної інженерії імені В.В. Поповського, Україна; email: viktor.solodilov@nure.ua

УДК 615.472 **DOI:10.30837/rt.2024.2.217.11**

В.В. СЕМЕНЕЦЬ, д-р техн. наук, О.В. ГРИГОР'ЄВ, канд. техн. наук

ДОСЛІДЖЕННЯ ПОКАЗНИКІВ КОЛІРНИХ ОБ'ЄКТІВ ЗА ДОПОМОГОЮ МІКРОКОНТРОЛЕРА STM32F407VG

Вступ

Прилади для вимірювання та визначення кольору використовуються протягом більше десяти років і нині знаходять широке застосування в електроніці, медицині, криміналістиці, сільському господарстві, космонавтиці та інших галузях. Важливість розробки та виробництва нових сучасних приладів та методів вимірювання кольору є актуальною задачею. Ці прилади та методи мають бути доступні, ефективні та прості у використанні [1].

Електронний колориметр – один із найпоширеніших інструментів вимірювання кольору, використовується в різних сферах через його переваги. До них відносяться такі як можливість кольорового контролю, простота в експлуатації, висока точність вимірювання та відносна дешевизна. Можливості електронного колориметра значно перевищують можливості інших кольорових вимірювальних приладів [2 – 6].

Розглянуто та проаналізовано широко використовувані моделі кольорів, а саме: модель RGB та субтрактивні моделі, зокрема CMY та CMYK. Досліджено характеристики колориметрів, які базуються на використанні моделей RGB, CMY та CMYK. Здійснено порівняльний аналіз переваг і недоліків моделей RGB та CMYK.

RGB модель

Це одна з найпопулярніших і широко використовуваних систем. Ця кольорова модель ґрунтується на трьох основних кольорах: червоний, зелений і синій. В системі RGB для представлення цих основних кольорів використовуються три монохроматичні довжини хвиль: $\lambda R = 700$ нм, $\lambda G = 546{,}1$ нм, $\lambda B = 435$ нм. Білий колір вважається основним стимулом у цій системі.

Ця кольорова модель є адитивною, що означає, що інтенсивність кольору зростає зі збільшенням яскравості окремих компонентів. Наприклад, змішуючи всі три кольори з максимальною інтенсивністю, отримаємо білий колір; навпаки, відсутність всіх кольорів призведе до чорного.

Модель CMY

Модель базується на принципі віднімання основних кольорів добавки з моделі RGB від білого, щоб утворити основні кольори. У цій концепції, кольори, які використовують біле світло для віднімання певних частин спектру, отримують назву "віднімаючих". Основні кольори в моделі CMY включають синій (результат віднімання червоного з білого), фуксин (іноді називаний фіолетовим) (результат віднімання зеленого з білого) та жовтий (результат віднімання синього з білого). Ці кольори утворюють друкарську тріаду та легко відтворюються на друкарських машинах.

Модель CMYK

Модель CMYK (Cyan–Magenta–Yellow–Key, де "Key" вказує на чорний колір) представляє собою значущий етап у розвитку кольорових моделей, що знайшли широке застосування у сучасному друку. Як і її попередниця CMY, модель CMYK є віднімаючою, враховуючи особливості сприйняття кольорів відповідно до принципів субтрактивної синтезу кольорів.

У моделі CMYK, символ чорного кольору позначається як "K," що походить від англійського "Key" (ключ). Відзначимо, що модель CMYK є емпіричною, орієнтованою на конкретне обладнання, відзначаючи відхід від теоретичних підходів моделей CMY та RGB.

Кожна точка на зображенні RGB розглядається як точка з різною яскравістю для кожного кольорового каналу, і для обчислення реальної яскравості використовується відповідна емпірична формула

$$Y = 0.2125 \cdot R + 0.7154 \cdot G + 0.0721 \cdot B$$

В апаратурі фотоелектричного колориметра автоматизовано виконується аналіз випромінювання та розрахунок кольорових координат за допомогою трьох селективних фотодетекторів. Спектральна чутливість цих фотодетекторів визначається за допомогою корекційних світлових фільтрів, які відповідають функціям додавання первинних кольорів. Кожен фотоприймач перетворює випромінювання своєї спектральної області в електричний струм. Цей процес гарантує пропорційність між вихідними електричними сигналами та вимірюваними кольоровими координатами.

Схема колориметра

Для подолання вказаних вище недоліків запропоновано використання найбільш економічного та простого у використанні колориметра та методу визначення кольорових характеристик, які відображені на рис. 1.

Ключовим елементом запропонованого для використання електронного колориметра є датчик кольору. Цей датчик надає три сигнали, пропорційні кольоровим координатам кожного випромінювання. Три фотоприймачі використовуються для перетворення світлової енергії в електричну енергію зі спектральною характеристикою, що розташована у видимій області спектру та відтворює одну з кольорових кривих складання.

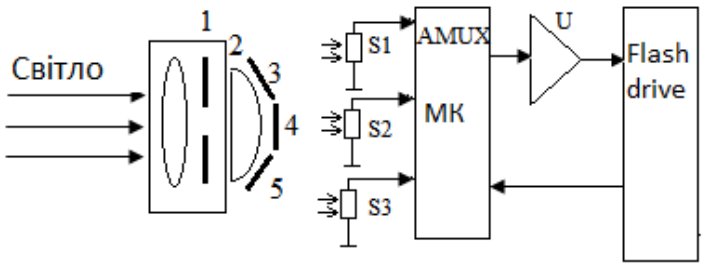


Рис. 1. Схема оптоелектронного колориметра

Сигнали, отримані від фотоприймачів, подаються на аналогові входи мікроконтролера(МК) STM32F407VG, з яких вони подаються на вбудований аналоговий мультиплексор (AMUX). З виходу нормалізуючого підсилювача сигнал передається на вхідний порт мікроконтролера (MCU). Мета нормалізуючого підсилювача полягає в тому, щоб привести вхідний сигнал в діапазон 0 ... 3,3В, що забезпечує достатній обсяг для оцифрування сигналу вбудованими 12-бітовими аналого-цифровими перетворювачами (АЦП) в мікроконтролерах [7, 8].

Мікропроцесор визначає необхідний канал, вводячи відповідний цифровий код на входи адресного мультиплексора. Вбудована програма управління пристроєм прошита всередині мікропроцесора. Програмне забезпечення мікропроцесора відповідає за ефективну роботу усього пристрою. Його функції включають управління перемиканням каналів аналогового мультиплексора, оцифрування нормалізованого сигналу від світлових датчиків, що надходять на вхід АЦП, калібрування сигналів від світлових датчиків, цифрову фільтрацію та усереднення результатів вимірювань, конвертацію вимірів освітленості до стандартного формату RGB та передачу результатів вимірювань та обчислень на дісплей або Flasl drive.

У мікроконтролерах серії STM32 вбудований 12-розрядний аналого-цифровий перетворювач. Його можливостей досить для обробки сигналів датчиків та дійсних сигналів. На його внутрішній вхід можна подавати аналогову напругу до 3,6 В з 16 зовнішніх входів чи від

вбудованих: датчика температури або джерела опорної напруги. При живленні контролера 3,3 В та роздільній здатності 12 біт діапазон напруги на вході від 0 В до 3,3 В буде перетворений у діапазон кодів від 0 до 4095. Максимальна частота перетворення 2,4 МГц.

У автоматичному режимі на аналого-цифровий перетворювач періодично подаються команди запуску перетворювання (рис. 2). Джерелом цих команд може бути, наприклад, подія від таймера.

Контролер DMA передає результат перетворення у масив в оперативній пам'яті. Якщо обрано декілька каналів вимірювання, то одна подія від таймера запускає послідовне перетворення у всіх обраних каналів. Кількість елементів масиву для зберігання результатів повинна бути пропорційна кількості каналів. Наприклад, якщо у нас два вхідних канали, то кількість елементів масиву для зберігання результатів повинна бути: 2, 4, 6 тощо.

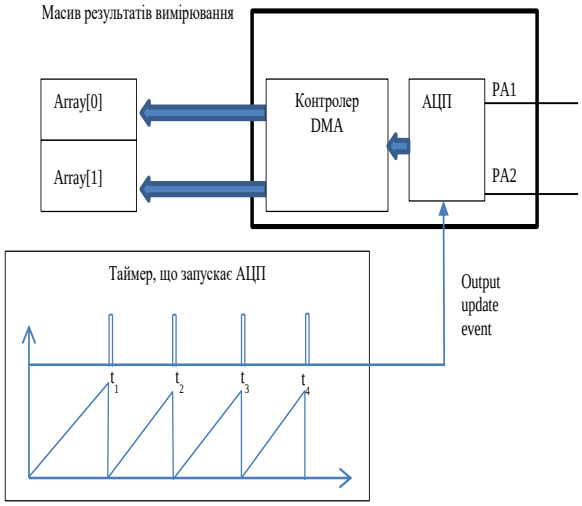


Рис. 2. Автоматичний режим роботи АЦП разом з контролером DMA

Налаштування АЦП у STM32CubeIDE

На вкладці Pinout&Configuration обираємо один з трьох АЦП (рис. 3). Далі треба обрати канали, на які будуть подаватись вхідні сигнали у вікні Mode. Як тільки обираємо один або декілька каналів, вони з'являються на зображенні мікроконтролера. Після цього налаштовуємо АЦП на вкладці Parameter Settings.

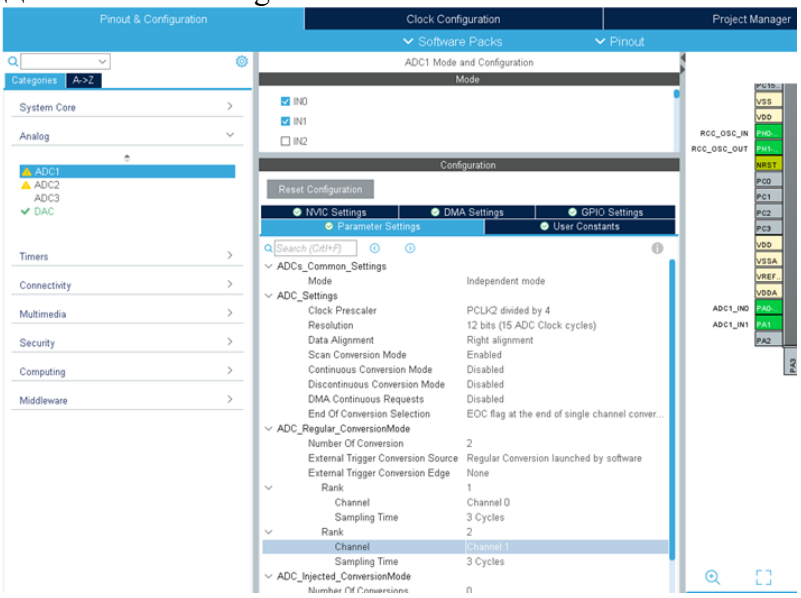


Рис. 3. Налаштування АЦП у STM32CubeIDE

Роздільна здатність АЦП (параметр Resolution) може дорівнювати: 12, 10, 8, 6 біт. Чим більше значення цього параметра тим точніше результат перетворення та більше відношення сигнал/завада, але менше швидкість перетворення. Якщо роздільна здатність дорівнює 12 біт, то отримуємо після перетворення значення коду у діапазоні від 0 до 4095, якщо роздільна здатність 10 біт, то діапазон кодів від 0 до 1023 тощо.

Орієнтація результату перетворення у вихідному регістрі (параметр Data Alignment) приймає значення: Right Alignment або Left Alignment, що відповідає розташуванню результату у молодших чи старших розрядах 16 розрядного вихідного регістру.

При перетворенні декількох каналів за допомогою параметра Scan Conversion Mode можна обрати автоматичне перемикання каналів. Тобто після перетворення сигналу у першому каналі аналоговий мультиплексом автоматично підключає наступний канал до входу АЦП і починається нове перетворення.

Параметр Continuous Conversion Mode дозволяє обрати режим безперервних перетворень.

Параметр DMA Continuous Request дозволяє використовувати канал DMA для передавання результатів перетворення у оперативну пам'ять, якщо цей канал вже створений та налаштований на вкладниці DMA settings.

Методика оцінки якості тестового зразка передбачає використання наступної процедури: тестовий зразок розміщується на столі, розташованому у фокальній площині приладу. Зразок піддається бічному або нижньому освітленню через прозорий стіл. Освітлення включає червоне, зелене та синє випромінювання. Вимірювання здійснюється для кожного кольору окремо, в будь-якій комбінації або для всіх кольорових джерел одночасно.

Експериментальні дослідження (табл. 1) з використанням розробленого електронного колориметра були проведені для аналізу вихідних сигналів вимірювальних каналів в залежності від яскравості випромінювання світла та різних положень фотодіодів відносно корпусу.

Таблиця 1

Результати вимірів

Номер	Червоний U_i, мВ	Синій U_i, мВ	Зелений U_i, мВ
1	148	90	95
2	149	93	99
3	148	90	98
4	149	91	97
5	148	92	96
6	150	92	95
7	150	91	99
8	149	91	98
9	151	90	97
10	151	93	97
11	148	92	95
12	149	90	95
13	148	91	96
14	150	93	96
15	151	93	99
16	149	92	99
17	148	92	98
18	149	91	98

Висновки

Існує три основних режими роботи колориметра: ручний, безперервний та автоматичний.

В першому режимі подається програмна команда для початку аналого-цифрового

перетворення. Коли перетворення закінчується виникають переривання від АЦП або від каналу DMA, що передає дані у оперативну пам'ять. Після цього можемо обробити отримані дані. Щоб отримати новий результат необхідно знову подати команду на початок перетворення.

У безперервному режимі перетворення ведуться з максимальною швидкістю. Як тільки завершується одне перетворення, починається наступне.

У автоматичному режимі на аналого-цифровий перетворювач періодично подаються команди від таймера запуску перетворювання. Контролер DMA передає результат перетворення у масив в оперативній пам’яті.

Розроблений цифровий колориметр дозволяє здійснювати експрес-контроль кольорових характеристик з високою метрологічною точністю і не потребує високого рівня кваліфікації обслуговуючого персоналу. Зазначений прилад відтворює сприйняття кольору людським зором, і його використання призначено як для осіб з нормальним, так і з ненормальним зором.

Великою перевагою цього пристрою є його мобільність, швидкість вимірювань і відсутність необхідності у спеціалізованій лабораторії та висококваліфікованих фахівцях.

Проведені повномасштабні вимірювання вихідної напруги на фотодіодах колориметра, що відповідають червоному, синьому та зеленому кольорам.

Список літератури:

1. Дудяк В. О. Природа кольору та його характеристики / В. О. Дудяк, Н. В. Занько, З. М. Сельменська. Львів : Укр. акад. друкарства, 2013. 208 с.

2. Кармазін В.В., В.В. Семенець. Курс загальної фізики. Київ : Кондор, 2009. 786 с.

3. Використання DICOM зображень в медицинських системах / М.Ю. Тимкович, О.Г. Аврунін, В.В. Семенець // Журнал НТУУ «КПІ» Техн. Електродинаміка. Силова електроніка та енергоефективність, (СЕЕ’2012)». 2012. С.178–183.

4. Печенюк Т. Кольорознавство. Київ : Грані-Т, 2009. 192 с.

5. Hahanov V.I., I.V. Hahanova. Design of digital systems by using VHDL language. Kharkov : KhNURE, 2003. 492c.

6. L.A. Vlasenko, A.G. Rutkas, A.A. Chikriy On the optimal impulse control in descriptor systems // Journal of automation and Information Sciences 51 (5). 2019.

7. Teaching microcontrollers and FPGAs in Quarantine from Coronavirus: Challenges and Prospects / O. Vorgul, I. Svyd, O. Zubkov, V.i Semenets. MC&FPGA, 2020.

8. Семенець В.В., Григорьєв О.В. Програмно-апаратний комплекс на базі мікроконтролера STM32F407VG для дослідження вібрацій акселерометром LIS3DSH // Радіотехніка. 2024. Вип. 216. С. 81–86.

Надійшла до редколегії 07.06.2024

Відомості про авторів:

Семенець Валерій Васильович – д-р техн. наук, проф., Харківський національний університет радіоелектроніки, професор кафедри біомедичної інженерії, Україна; e-mail: valery.semenets@nure.ua; ORCID: https://orcid.org/0000-0001-8969-2143

Григор’єв Олександр Вікторович – канд. техн. наук, доц., Харківський національний університет радіоелектроніки; Україна; e-mail: oleksandr.hryhoryev@nure.ua; ORCID: https://orcid.org/0000-0001-6467-7983

УДК 621.396 **DOI:10.30837/rt.2024.2.217.12**

І.М. МИЦЕНКО, д-р фіз.-мат. наук, Ю.О. ПЕДЕНКО, канд. техн. наук,
О.М. РОЄНКО, канд. фіз.-мат. наук

ПРО МОЖЛИВІСТЬ ЗАХИСТУ БПЛА ВІД ПРИДУШЕННЯ СИГНАЛІВ УПРАВЛІННЯ

Вступ

У теперішній час безпілотні літальні апарати (БПЛА) широко застосовують у найрізноманітніших секторах як народного господарства, так у військових цілях. У [1] представлено досить вичерпний огляд різних платформ БПЛА відповідно до різних класифікацій, надано опис новітніх технологій, що використовуються в БПЛА. Розвиток БПЛА надав широкий шлях щодо комерціалізації. БПЛА вже використовуються в сільськогосподарській промисловості для моніторингу сільськогосподарських угідь, аналізу зразків ґрунту та навіть для пасіння худоби. У майбутньому він може розширитися ще більше, тоді як попит постійно зростає. БПЛА також використовуються в пошуково-рятувальних роботах, щоб рятувати людей із ситуацій, що загрожують життю, і щоразу ця технологія доводить свою корисність як ніколи. У найближчому майбутньому не дивно, що БПЛА знайдуть своє застосування в роздрібній торгівлі, транспорті, розвагах, охороні житла та навіть у будівництві з використанням 3D-принтерів. Зрозуміло, що майбутні можливості для БПЛА величезні. Крім того, з розвитком технологій стало легше виробляти безпілотники та керувати ними. У сучасну епоху, якщо БПЛА об’єднати з технологією смартфонів, це призведе до створення безпечних, надійних інструментів і функцій [2 – 5]. Також слід нагадати, що БПЛА десятиліттями використовувалися збройними силами різних країн для авіарозвідки, управління вогнем і цілевказівки, нанесення ударів по наземних та інших цілях та інше.

Одночасно зі зростанням можливостей БПЛА, розвивалися методи і засоби, що їм протидіють. У [6] підкреслено, що виявлення, класифікація, ідентифікація, відстеження та подолання низьких, повільних і малих повітряних загроз є серйозною проблемою для існуючих сенсорних систем моніторингу. Системи так званого протидії безпілотним літальним апаратам першого покоління часто покладаються на виявлення каналу передачі даних від оператора до БПЛА, що забезпечує обмежені можливості проти поточних загроз. Однак цей спосіб виявлення БПЛА є проблемою, коли оператори маніпулюють стандартними каналами передачі даних, і він взагалі не працюватиме проти поточних і майбутніх автономних БПЛА. Інші сучасні методи виявлення та нейтралізації БПЛА включають, наприклад, поєднання радара з оптичними датчиками. Ці системи не завжди надійні, можуть генерувати велику кількість хибних тривог і часто вимагають великої кількості обслуговуючого персоналу [7, 8].

Найдоступнішим методом боротьби з БПЛА є придушення радіосигналів, які використовуються для його управління. У результаті БПЛА втрачає зв'язок з оператором і не може продовжити рух у робочому режимі, що часто призводить до його краху. Тому розробка методів і засобів боротьби з придушенням сигналів керування не втрачає своєї актуальності та стає все більш необхідною. У [9] наведено та проаналізовано методи захисту каналу передачі БПЛА, що є першочерговим завданням для забезпечення протидії атакам на БПЛА. При цьому треба враховувати, що навіть використання найбільш захищених від впливу умисних завад видів модуляції з розширенням спектра не гарантує захисту такого каналу. Досліджено структурно-функціональні методи побудови бездротової захищеної системи каналів зв'язку БПЛА. Запропоновано архітектурне рішення із використанням двох каналів у різних діапазонах частот для каналу управління БПЛА. Подано схематичну структуру організації такого каналу зв'язку. Дано вираз запасу міцності каналу зв'язку проти конкретної навмисної завади. Показано, що запропоноване архітектурне рішення матиме аналогічний ефект при впливі на канал зв'язку структурованих завад. У разі впливу імітаційної завади ситуація буде неодно-

значною, тому дуже важливо правильно визначити канал, на який впливає навмисна завада [10 – 12].

Одним із методів боротьби з придушенням радіосигналів, що керують БПЛА, є (у разі виявлення сигналів завад) зміна робочої частоти сигналу керування БПЛА на іншу, що дає змогу БПЛА продовжувати рух і не втрачати працездатність. Цей метод і покладено в основу пропонованого в статті рішення. Метою даної роботи є розробка схеми автономного приймального пристрою (блоку захисту), що підключається до входу приймача БПЛА та здійснює його захист від радіосигналів завад у робочому режимі (рис. 1).

Рис. 1. Підключення блока захисту до приймача-передавача БПЛА

За відсутності сигналів завад на штатній частоті БПЛА $f_Ш$ блок захисту, приймає сигнал керування оператора і передає його на вхід приймача БПЛА, який рухається в робочому режимі і виконує поставлені перед ним завдання.

У разі появи радіосигналів завад на штатній частоті, тобто на частоті $f_З = f_Ш$ самостійно їх виявляє і за допомогою передавача БПЛА видає команду оператору на зміну робочої частоти $f_Ш$ на додаткову частоту $f_Д$. При цьому вихідна частота блоку захисту залишається постійною і дорівнює штатної частоті БПЛА $f_Ш$. Це дає змогу не перебудовувати приймач БПЛА, що істотно спрощує застосування блока захисту. Крім цього, при отриманні інформації про наявність радіосигналів завад блок керування БПЛА також автоматично переходить на додаткову робочу частоту $f_Д$. БПЛА продовжує рух і виконання поставлених перед ним завдань, але використовує додаткову робочу частоту.

Таким чином, до завдань блоку захисту входить:

- постійний контроль з метою виявлення наявності радіосигналів завад на штатній робочій частоті $f_Ш$;
- за відсутності радіосигналів завад приймання радіосигналів керування від оператора на штатній частоті $f_Ш$ та передавання їх на вхід приймача БПЛА;
- у разі виявлення радіосигналів завад на штатній частоті $f_Ш$ передача інформації про їх наявність оператору, що призводить до автоматичної зміни частоти передавача керування на додаткову частоту $f_Д$;
- також у разі виявлення радіосигналів завад здійснюється автоматичне перемикання блока захисту на приймання керуючих сигналів на додатковій частоті $f_Д$. Радіосигнали додаткової частоти у блоці захисту перетворюються таким чином, щоб його вихідний сигнал мав частоту, що дорівнює штатній частоті БПЛА $f_Ш$. Після перетворення вихідний сигнал блока захисту подається на вхід приймача БПЛА.

Для можливості керування БПЛА та одночасного виявлення сигналів завад у пропонованому методі використовують два режими роботи, що чергуються (див. рис. 2).

Під час першого режиму $T_{роб}$ здійснюється штатне керування БПЛА на робочій частоті $f_Ш$. Другий режим $T_{вияв}$ призначений для знаходження радіосигналу завади на штатній частоті $f_Ш$. В цьому режимі сигнал керування оператора не випромінюють, а блок захисту

БПЛА контролює наявність радіосигналів на частоті $f_{\text{ш}}$. Виявлення сигналу на частоті $f_{\text{ш}}$ в другому режимі свідчить про наявність завади.

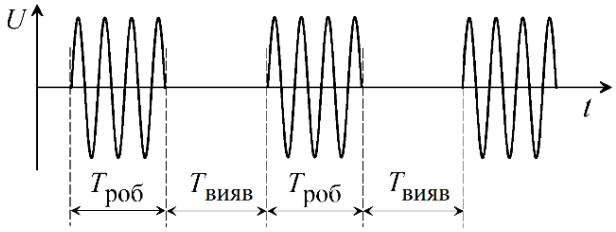


Рис. 2. Керуючий сигнал БПЛА

У період часу $T_{\text{роб}}$ сигнал управління випромінюється передавачем оператора на основній частоті $f_{\text{ш}}$ і управляє БПЛА, який виконує своє завдання, потім у період часу $T_{\text{вияв}}$ сигнал управління не випромінюється, що дає можливість блоку захисту виявляти сигнали завад на робочій частоті $f_{\text{ш}}$. При цьому, якщо сигнал завади блоком захисту не виявлено, БПЛА продовжуватиме свою роботу на штатній частоті $f_{\text{ш}}$. У разі виявлення завади на частоті $f_{\text{ш}}$ блок захисту автоматично змінює робочу частоту на додаткову частоту $f_{\text{д}}$, при цьому БПЛА зберігає можливість виконання свого завдання. Тривалість режиму роботи $T_{\text{роб}}$ і знаходження $T_{\text{вияв}}$ обирають таким чином, щоб не порушувався робочий режим БПЛА з урахуванням його швидкості польоту, швидкості обробки сигналів та інших можливостей.

Таким чином, блок захисту має являти собою простий і недорогий пристрій, який би виявляв сигнали завади й автоматично перемикав керування БПЛА на додаткову робочу частоту $f_{\text{д}}$. При цьому частота вихідного сигналу блок захисту залишається незмінною і дорівнює штатній робочій частоті $f_{\text{ш}}$.

На рис. 3 наведено функціональну схему блока захисту БПЛА.

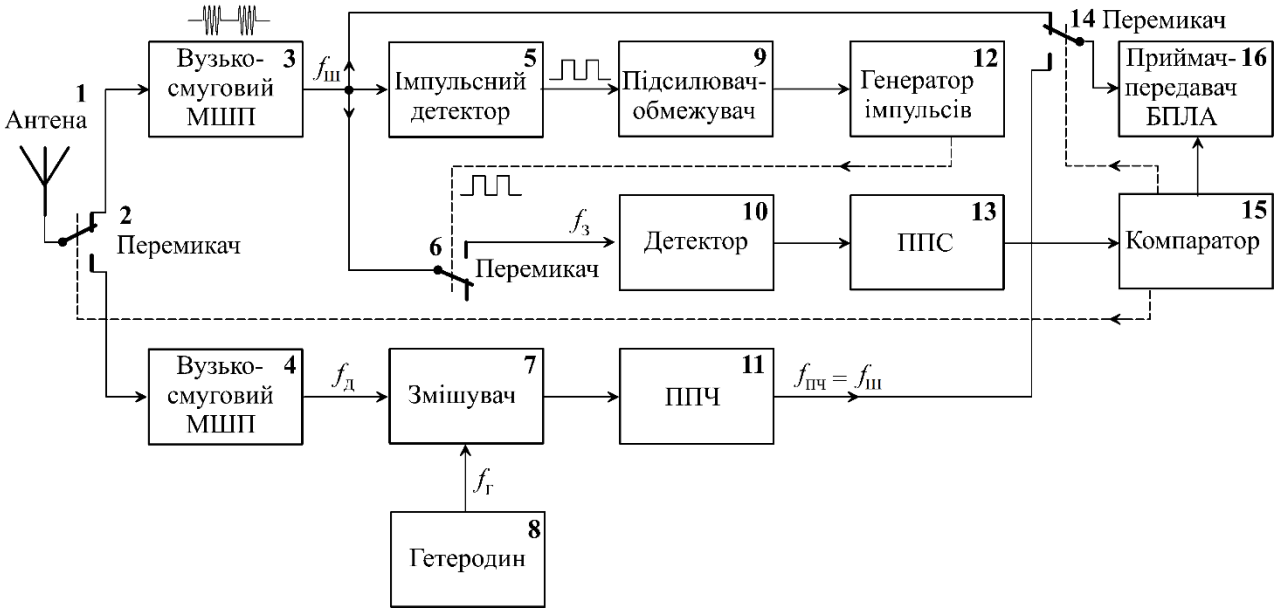


Рис. 3. Функціональна схема блоку захисту БПЛА

Працює блок захисту таким чином. У робочому режимі (період часу $T_{роб}$) сигнал управління на частоті $f_{Ш}$ від оператора БПЛА приймається антеною 1, що має робочу смугу частот, достатню для приймання сигналів як на штатній частоті $f_{Ш}$, так і на додатковій частоті $f_{Д}$, для цього може бути використана зокрема логоперіодична антена [13]. У разі великого розносу робочих частот $f_{Ш}$ и $f_{Д}$ слід застосовувати дводіапазонну антену, наприклад [14].

Далі прийнятий сигнал за допомогою антенного перемикача 2 надходить на вхід малошумливого підсилювача (МШП) 3. Резонансний малошумливий підсилювач 3 налаштований на основну (штатну) робочу частоту $f_{Ш}$. Смуга амплітудно-частотної характеристики має бути не вужчою за спектр сигналу керування, що приймається.

Прийнятий сигнал керування посилюється підсилювачем 3 до величини, необхідної для роботи виявника сигналів завад і синхронізації всього блока захисту загалом. Для цього його подають на вхід імпульсного детектора 5, де відбувається виділення його огинаючої, а також на вхід комутатора 6 пристрою виявлення завад.

Крім цього, вихідний сигнал підсилювача 3 за допомогою комутатора 14 подається безпосередньо на вхід приймача БПЛА, який виконує своє завдання і перебуває в робочому режимі.

Керування блоком захисту здійснюється синхронізатором, який складається з імпульсного детектора 5, підсилювача-обмежувача 9 і генератора імпульсів 12. Вихідний сигнал імпульсного детектора 5 підсилюється та обмежується підсилювачем-обмежувачем 9, який обмежує вершини сигналів і прибирає небажані флуктуації амплітуд сигналу керування БПЛА.

Далі вихідний сигнал підсилювача-обмежувача 9 подається на вхід синхронізації генератора імпульсів 12 і синхронізує генерацію його імпульсів з імпульсами $T_{роб}$ вхідного сигналу оператора. Генератор імпульсів 12 працює в режимі очікування і під час приходу імпульсу запуску з виходу підсилювача-обмежувача 9 генерує імпульс тривалістю $T_{роб}$ позитивної полярності, що подається на схему виявника сигналів завад, а саме на керуючий вхід комутатора, та тримає його в розімкнутому стані. Коли дія цього імпульсу закінчується і настає час знаходження $T_{вияв}$, перемикач 6 підключає вхід схеми виявлення до виходу малошумливого підсилювача 3.

Схема виявлення складається з вхідного перемикача 6, детектора сигналу завад 10, підсилювача постійного струму (ППС) 13 і компаратора 15. Він працює таким чином. Під час $T_{роб}$, коли випромінюється сигнал керування, схема виявлення відключена від вхідних пристроїв блока захисту і сигнали керування на неї не надходять. Потім, у період виявлення завад $T_{вияв}$, коли відсутнє випромінювання сигналів керування, її вхід за допомогою перемикача 6 підключають до вхідних пристроїв блока захисту і вона контролює наявність сигналів завад на штатній робочій частоті $f_{Ш}$. За відсутності сигналів завад у період $T_{вияв}$ вихідний сигнал схеми виявлення дорівнює нулю, а компаратор 15 не спрацьовує та залишає антенний перемикач 2, перемикач 14 у початковому положенні. При цьому БПЛА продовжує рух із використанням штатної частоти $f_{Ш}$.

У разі потрапляння БПЛА в зону дії сигналів завад на частоті $f_{Ш}$ в період $T_{вияв}$, коли відсутній сигнал керування від оператора, блок захисту їх виявляє й автоматично перемикає блок захисту на додаткову частоту $f_{Д}$. Крім цього каналом зв'язку передає інформацію про наявність сигналів завад на передавальний пристрій оператора, який також автоматично перемикається на додаткову частоту $f_{Д}$. У результаті, за наявності сигналів завад на частоті $f_{Ш}$ схема виявлення і захисту БПЛА працює таким чином. У період виявлення $T_{вияв}$ сигнал

завади за допомогою комутаторів надходить на вхід детектора 10, де детектується і далі посилюється підсилювачем постійного струму ППС 13.

Вихідний сигнал підсилювача 13 подається на керуючий вхід компаратора 15, який спрацьовує і змінює знак вихідного сигналу на протилежний. Спрацьовування компаратора 15 призводить до спрацьовування антенного перемикача 2, який антену 1 відключає від входу малошумливого підсилювача 3 і підключає її до входу малошумливого підсилювача 4 додаткової частоти $f_д$. Крім цього, спрацьовування компаратора 15 призводить до спрацьовування комутатора 14, який відключає від входу БПЛА сигнал керування на робочій частоті $f_ш$ і підключає вхід БПЛА до виходу приймача додаткової частоти $f_д$. Приймач додаткової частоти $f_д$ складається з малошумливого резонансного підсилювача 4, змішувача 7, гетеродина 8 і підсилювача проміжної частоти ППЧ 11 та працює таким чином. Після спрацьовування компаратора 15 антенний перемикач 2 під'єднує антену 1 до входу малошумливого резонансного підсилювача 4 сигналів керування на додатковій робочій частоті $f_д$, який їх вибірково підсилює і подає на вхід перетворювача частоти, де вони перетворюються так, щоб вихідна проміжна частота $f_{пч}$ дорівнювала основній (штатній) робочій частоті БПЛА $f_ш$:

$$f_{пч} = f_г - f_д = f_ш, \text{ якщо } f_г > f_д, \quad (1)$$

$$f_{пч} = f_д - f_г = f_ш, \text{ якщо } f_г < f_д, \quad (2)$$

де $f_г$ – частота гетеродина 8,

Слід мати на увазі, що під час модуляції сигналу керуючого БПЛА, де відіграють роль фазові зсуви, кращим є вибір $f_г > f_д$ (1), оскільки зберігається знак фазових зсувів. У разі $f_г < f_д$ (2), знак змінюється на протилежний. Для цього випадку сигнал на виході підсилювача проміжної частоти необхідно забезпечити кут зсуву фаз, що дорівнює 180°. У першому випадку в цьому немає необхідності [15].

Крім цього, вихідний сигнал порогового пристрою 15 у разі виявлення завади і спрацьовування подається на передавач БПЛА, що здійснює зв'язок з оператором, і блок управління оператора автоматично переходить на додаткову частоту $f_д$.

Висновки

Пропонований метод і схема пристрою, що його реалізує (блок захисту), для боротьби з придушенням радіосигналів, що керують БПЛА, дають можливість:

- вести постійний контроль і виявлення радіосигналів завади в разі їх появи;
- за відсутності радіосигналів завад здійснювати приймання радіосигналів керування від оператора на штатній частоті $f_ш$ і передавати їх на вхід приймача БПЛА;
- у разі виявлення радіосигналів завад на основній частоті f штатній частоті $f_ш$ перемикати приймач блока захисту на додаткову частоту $f_д$, водночас частота вихідного сигналу блока захисту після перетворення залишається такою самою, як частота $f_ш$, що дає змогу зберегти налаштування приймача БПЛА;
- каналом зв'язку з оператором надсилати інформацію про наявність завад і автоматично перемикати блок керування БПЛА на додаткову робочу частоту.

Список літератури:

1. Chaurasia R., Mohindru V. Unmanned Aerial Vehicle (UAV): A Comprehensive Survey. DOI: https://doi.org/10.1002/9781119769170.ch1
2. Drones Importance and Usage in Real Estate. https://www.ifsec.events/india/visit/news-and-updates/drones-importance-and-usage-real-estate. Accessed 31 March 2021.
3. Cai G., Dias J., Seneviratne L. A Survey of Small-Scale Unmanned Aerial Vehicles: Recent Advances and Future Development Trends. Unmanned Syst., 2, 2, 175–199, 2014.
4. Pham H., Smolka S.A., Stoller S.D., Phan D., Yang J. A survey on unmanned aerial vehicle collision avoidance systems. CoRR, abs/1508.07723, 2015.
5. Hackney C. and Alexander C., 2.1.7. Unmanned Aerial Vehicles (UAVs) and their application in geomorphic mapping // Geomorphological Techniques. L. Clarke, and J. M. Nield, (Eds.), British Society for Geomorphology, London, GB, 2015.
6. Dominicus J. New Generation of Counter UAS Systems to Defeat of Low Slow and Small (LSS) Air Threats. https://apps.dtic.mil/sti/pdfs/AD1152139.pdf
7. Willis M., et al. A Comprehensive Approach to Countering Unmanned Aircraft Systems // Joint Air Power Competence Centre, (2021).
8. Holland Michel Arthur. Counter-Drone Systems. 2 nd Edition, Center for the Study of the Drone at Bard College, https://dronecenter.bard.edu/projects/counter-drone-systems-project/counter-drone-systems-2nd-edition, (2019).
9. Kaidenko M. M., Kravchuk S. O. Protection against the effect of different classes of attacks on UAV control channels // Information and telecommunication sciences. 2022. Vol. 13, №1. P. 35–43.
10. B. Sklar. Digital Communications: Fundamentals and Applications, Second Edition, Prentice-Hall, Upper Saddle River, NJ, 2001.
11. A. Wiesel, J. Goldberg, H. Messer-Yaron. SNR estimation in time-varying fading channels // IEEE Transactions on Communications, vol. 54, no. 5, pp. 841-848, May 2006, http://dx.doi.org/10.1109/TCOMM.2006.873995.
12. Z. Sun, X. Gong, F. Lu. A non-data-aided SNR estimator based on maximum likelihood method for communication between orbiters // JWireless Com Network, 123 (2020). https://doi.org/10.1186/s13638-020-01730-4
13. Optimization of a printed log-periodic antenna position on a UAV / Venkata Reddy Kandregula, Pavlos Lazaridis, Zaharias Zaharis, et al. // TechRxiv. October 26, 2023. DOI:10.36227/techrxiv.24427117.v1
14. Akhter Z, Bilal R. M., Shamim A. A. Dual mode, Thin and Wideband MIMO Antenna System for Seamless Integration on UAV // IEEE Open Journal of Antennas and Propagation. 2021. Vol. 2, September. P. 991–1000. DOI: 10.1109/OJAP.2021.3115025
15. Бова М. Т., Гойжевський В. О., Маєвський С. М., Молебний В. В. Вимірювання різниці фаз у радіоелектроніці. Київ : Вища шк., 1972. 232 с.

Надійшла до редколегії 29.05.2024

Відомості про авторів:

Миценко Ігор Михайлович – д-р фіз.-мат. наук, Інститут радіофізики та електроніки імені О.Я. Усикова НАН України, ст. науковий. співробітник; Україна; email: igor.mytsenko@gmail.com: ORCID: https://orcid.org/0000-0001-6598-6809

Педенко Юрій Олександрович – канд. техн. наук, Інститут радіофізики та електроніки імені О.Я. Усикова НАН України, ст. науковий. співробітник; Україна; email: yuriy.pedenko@gmail.com; ORCID: https://orcid.org/0009-0006-8752-9581

Роєнко Олександр Миколайович – канд. фіз.-мат. наук, Інститут радіофізики та електроніки імені О.Я. Усикова НАН України, зав. відділом, Україна; email: alexnikrnk@gmail.com; ORCID: https://orcid.org/0000-0001-9632-527X

УДК 621.357 **DOI:10.30837/rt.2024.2.217.13**

В.М. БОРЩОВ, д-р техн. наук, О.М. ЛІСТРАТЕНКО, канд. техн. наук,
М.І. СЛІПЧЕНКО, д-р фіз.-мат. наук, М.А. ПРОЦЕНКО, канд. техн. наук,
І.Т. ТИМЧУК, канд. техн. наук, О.В. КРАВЧЕНКО, І.В. БОРЩОВ

ДОСЛІДЖЕННЯ ТЕПЛОВИХ ВЛАСТИВОСТЕЙ ЕЛЕКТРОННИХ МОДУЛІВ НА КОМБІНОВАНИХ ПЛАТАХ З ПОЛІІМІДНИМИ ДІЕЛЕКТРИКАМИ

Вступ

Застосування в комбінованих платах на металевих жорстких основах у якості діелектриків поліімідних (ПІ) плівок дає змогу виділити такі плати за конструкцією в особливий ряд, оскільки в них використовують ПІ діелектрики порівняно малої товщини, близько 0,02 – 0,025 мм проти порядку 0,1 мм у діелектриків, які застосовують в існуючих серійних платах на металевих основах. Хоча промислові ПІ плівки мають низькі значення теплопровідності близько 0,12 – 0,14 Вт/(м·К), проте їхня невелика товщина в платах забезпечує доволі малий тепловий опір тепловідвідної системи загалом. При цьому ПІ плівки, не дивлячись на малу товщину, мають високу електричну міцність (до 160 кВ/мм та більше) порівняно з іншими типами плат на алюмінієвих основах. А використання в якості композиційних, зокрема ПІ плівок з підвищеною теплопровідністю, дає змогу ще більше зменшити сумарний тепловий опір друкованих плат на металевих основах з тонкими ПІ діелектриками [1 – 3].

Серійні тонкі термозварювальні поліімід-фторопластові плівки (ПМФ), в тому числі теплопровідні, також можуть бути успішно застосовані для створення комбінованих плат на металевих жорстких основах, що вимагають покращених фізико-механічних та теплових властивостей матеріалів, стійкості до впливу температури та високої надійності за збереженням вже досягнутих для поліімідів інших функціональних властивостей високого рівня. Такий підхід дозволяє суттєво розширити інноваційні можливості нових електронних модулів і друкованих вузлів, що розробляються, практично для всіх областей спеціального приладобудування, у тому числі із застосуванням Chip-on-board (COB) і Chip-on-flex (COF) технологій складання [4 – 7].

Метою цієї роботи було побудова та теоретичні дослідження теплових моделей електронних модулів з підвищеною теплопровідністю на основі комбінованих плат з використанням промислових термозварювальних поліімід-фторопластовими плівок, у тому числі з теплопровідністю 0,12 – 0,46 Вт/м·К, а також з удосконаленими авторами лакофольговими діелектриками з теплопровідністю ПІ шарів порядку 4,0 – 4,5 Вт/(м·К). Проведення експериментальних досліджень ефективності відводу тепла від напівпровідникових пристроїв у тестових структурах якості електронних модулів на основі різних типів комбінованих плат з поліімідними діелектриками.

1. Предмет та методи дослідження

1.1. Теоретичні дослідження теплових властивостей електронних модулів на комбінованих платах з поліімідними діелектриками

Теоретичні дослідження теплових властивостей електронних модулів на комбінованих платах з ПІ діелектриками полягають у побудові та дослідженні їх теплових моделей. Чим нижчий тепловий опір, тим більше і швидше відведення тепла. Передача тепла з одного місця (наприклад, від напівпровідникового чипу) в інше (навколишнє середовище) визначається товщиною шарів і тепловим опором матеріалів, які застосовують, а також площею їхнього дотику (чим більша площа дотику, тим більшу кількість тепла може бути передано). Оскіль-

ки основною метою використання комбінованих металевих друкованих плат у радіоелектронній апаратурі є покращення теплопередачі від електронних тепло навантажених компонентів до системи забезпечення теплового режиму, в якості основного критерію при розрахунках доцільно розглядати тепловий опір у системі «напівпровідникова структура – зворотний бік друкованої плати». При ідентичності таких величин, як площа плати, товщина металевої фольги, товщина та властивості металевої основи, а також товщина шару припою або клейового з'єднання між напівпровідниковим чипом та контактною площиною на металевій фользі з друкованими провідниками, товщина ПІ діелектрика та його теплопровідність будуть визначальними для значень сумарного теплового опору комбінованих плат.

На рис. 1 зображено типова структурна схема електронного модулю підвищеної потужності на комбінованій теплопровідній платі з ПІ діелектриком. У структурній схемі електронного модулю одиничне джерело тепла встановлено за допомогою паяння на контактні площини верхнього шару плати, який виконаний з мідної фольги. У комбінованій платі застосовано плоску алюмінієву теплопровідну основу та діелектричний поліімідний шар, який з'єднаний із шаром мідної фольги та теплопровідною основою за допомогою тонких діелектричних фторполімерних термозварюваних плавких покриттів або за допомогою адгезивних шарів. При цьому прийняті умови, що все тепло, що виділяється від одиничного джерела тепла (напівпровідникового чипу) без втрат передається шару припою, який розташовано на високо теплопровідній мідній контактній площині верхнього комутаційного шару комбінованої плати.

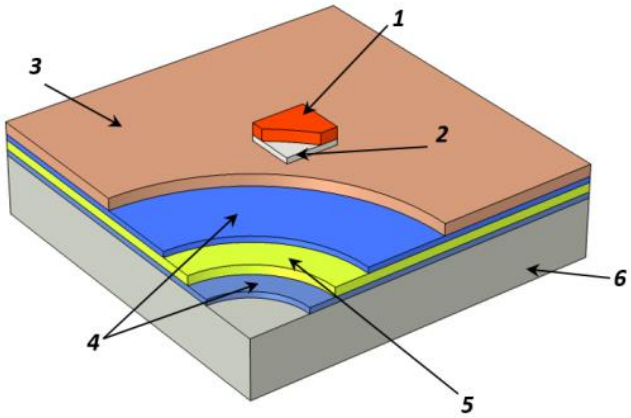


Рис. 1. Типова структурна схема електронного модулю
1 – джерело тепла; 2 – шар припою; 3 – шар фольги з міді; 4 – шари термозварювальних покриттів або адгезиву; 5 – шар поліміду; 6 – алюмінієва основа

1.2. Теплове моделювання електронних модулів на комбінованих платах з поліімідними діелектриками

Для теоретичних досліджень теплових властивостей електронних модулів на комбінованих платах з поліімідними діелектриками було відібрано чотири типи плат. В тому числі: плата на основі термозварювальної ПМФ плівки Kapton® 120FN 616 компанії DuPont (США) зі стандартною ПІ плівкою-основою DuPont™ Kapton® HN товщиною 25 мкм з мінімальною теплопровідністю 0,12 Вт/(м·К) з фторполімерними двосторонніми покриттями Teflon® FEP товщиною 2,5 мкм кожне з теплопровідністю 0,20 Вт/(м·К) (Модель 1) [4, 2], плата на основі термозварювальної ПМФ плівки KYMIDE KYPIFEP 9198 (FHF) компанії Suzhou Kying Industrial Materials Co.Ltd (Китай) зі стандартною ПІ плівкою-основою товщиною 30 мкм з мінімальною теплопровідністю 0,12 Вт/(м·К) з фторполімерними двосторонніми покриттями FEP товщиною 10 мкм кожне з теплопроводностью 0,20 Вт/(м·К) (Модель 2) [5] та плата на основі термозварювальної ПМФ плівки DuPont Kapton®120FMT616 компанії DuPont з теплопровідною ПІ плівкою-основою Kapton® MT товщиною 25 мкм з теплопровідністю

0,46 Вт/(м·К) та фторполімерними двосторонніми покриттями Teflon® FEP товщиною 2,5 мкм кожне з теплопровідністю 0,20 Вт/(м·К) (Модель 3) [6]. А також комбінована плата на основі розробленого авторами інноваційного одностороннього лакофольгового мідь-поліімідного діелектрику з поліімідною композиційною плівкою з наповнювачем із суміші високотеплопровідних високодисперсних (8 мкм) та ультрадисперсних (0,4 мкм) порошків білого нітриду алюмінію (AlN) товщиною 25 – 30 мкм з розрахунковою теплопровідністю порядку 4,0 – 4,5 Вт/(м·К). Приклеювання безадгезивного мідь-поліімідного лакофольгового діелектрику до алюмінієвої основи плати здійснюється за допомогою шару однокомпонентного вологозатвердженого теплопровідного полідиметилсилоксанового адгезивного матеріалу Kafuter K-5204K компанії Guangdong Hengda New Materials Technology Co., Ltd. (KAFUTER) (Китай) товщиною 25 мкм з теплопровідністю 1,6 Вт/(м·К) (Модель 4) [8, 9].

Теоретичне моделювання розподілу тепла та ефективності його передачі від джерела тепла до металевої основи комбінованих плат із вибраними типами ПІ плівок у досліджуваних моделях електронних модулів проводилося за допомогою програмного комплексу COMSOL MULTIPHYSICS. У моделях електронних модулів досліджувалися теплові властивості плат з розмірами 35 × 35 мм на алюмінієвих основах товщиною 1000 мкм з теплопровідністю ~ 238 Вт/(м·К). У типовій структурі електронних модулів під час моделювання передбачено застосування мідної фольги товщиною 100 мкм з теплопровідністю ~ 400 Вт/(м·К) у верхньому шарі комбінованих друкованих плат. Одиничне джерело тепла з площею теплового контакту його основи 4,29 мм^2 з потужністю 2 Вт встановлено на мідну контактну площину верхнього шару плати за допомогою паяльної пасти з теплопровідністю 85 Вт/(м·К) із товщиною шару припою 30 мкм. При цьому в моделях досліджувалися теплові властивості плат для відносно сприятливих умов експлуатації модулів при температурі навколишнього середовища Та = 25 °С з конвекційним коефіцієнтом теплопередачі від плоских алюмінієвих основ h = 22 Вт/м^2 К та для відносно несприятливих умов експлуатації при Та = 45 °С з конвекційним коефіцієнтом теплопередачі від плоских алюмінієвих основ h = 17 Вт/м^2 К.

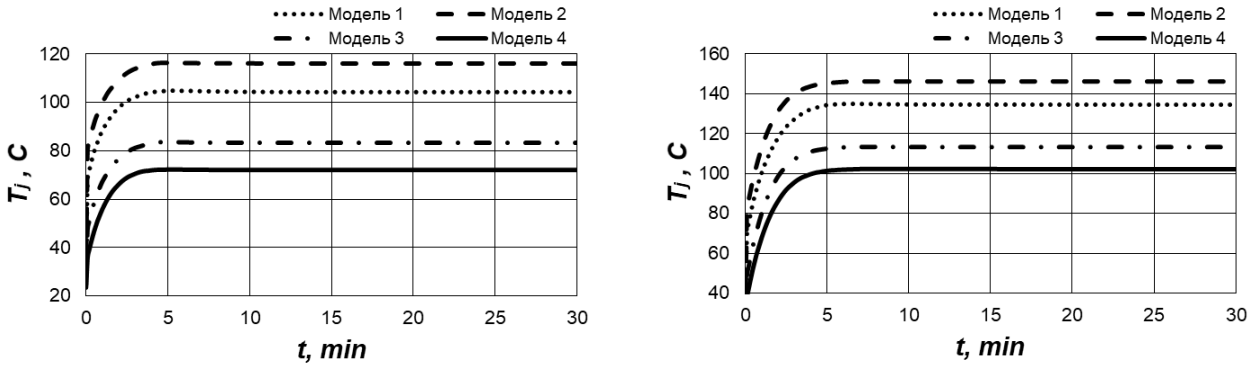
У табл. 1 наведено результати розрахунків сумарних теплових опорів комбінованих плат із шарами припоїв для теплових моделей електронних модулів. А також результати розрахунків температур T_j у шарах припоїв під джерелом тепла (в області джерела тепла) у комбінованих платах при стабільному стані теплового розподілу за сприятливих та несприятливих умов експлуатації модулів.

Таблиця 1

Результати розрахунків сумарних теплових опорів комбінованих плат і температур на платах у шарах припою під джерелами тепла при стабільному стані теплового розподілу за сприятливих і несприятливих умов експлуатації модулів

№ п/п	Варіанти теплових моделей	Сумарний тепловий опір комбінованих плат $R_к$, °С/Вт	Температура на платі в області джерела тепла, T_j, °С	
			При h = 22 Вт/м^2·К, Та = 25 °С	При h = 17 Вт/м^2·К, Та = 45 °С
1	Модель 1	0,276	104,1	134,4
2	Модель 2	0,372	116,1	146,3
3	Модель 3	0,151	83,2	113,4
4	Модель 4	0,109	72,1	102,2

На рис. 2 представлено залежності температур нагріву на платах в області джерела тепла від часу при переході в стабільний температурний стан за сприятливих та несприятливих умов експлуатації модулів.



а *б*

Рис. 2. Залежність температур нагріву на платах в області джерела тепла T_j від часу при переході в стабільний температурний стан за сприятливих (*а*) та несприятливих (*б*) умов експлуатації модулів

Із отриманих розрахункових результатів, які представлено у табл. 1 та на рис. 2, можна зробити висновок, що найменші значення сумарних теплових опорів комбінованих плат і температур на платах у області джерел тепла під час переходу в стабільний температурний стан було отримано для електронних модулів у Моделях 3 і 4. При цьому в типовій структурній схемі електронного модулю в Моделі 3 було досягнуто значення сумарного теплового опору комбінованої плати на основі термозварювальної ПМФ плівки Kapton®120FMT616 компанії DuPont порядку 0,15 ºС/Вт. У типовій структурній схемі електронного модулю в Моделі 4 було досягнуто значення сумарного теплового опору комбінованої плати на основі розробленого авторами композиційного одностороннього лакофольгового мідь-поліімідного діелектрику порядку 0,11 ºС/Вт. У Моделі 3 температура в області джерела тепла потужністю 2 Вт досягала 83,2 та 113,4 ºС після 20 хвилин, переходячи в стійкий стан відповідно за сприятливих умов та за несприятливих умов експлуатації. У Моделі 4 температура в області джерела тепла потужністю 2 Вт досягала 72,1 ºС та 102,2 ºС після 20 хвилин, переходячи в стійкий стан відповідно за сприятливих і за несприятливих умов експлуатації.

У Моделях 1 та 2 теплові характеристики електронних модулів були істотно гірше порівняно з Моделями 3 та 4. У Моделі 1 було досягнуто значення сумарного теплового опору комбінованої плати на основі термозварювальної ПМФ плівки Kapton® 120FN 616 компанії DuPont порядку 0,28 ºС/Вт, а у Моделі 2 значення сумарного теплового опору комбінованої плати на основі термозварювальної ПМФ плівки KYMIDE KYPIFEP 9198 (FHF) компанії Suzhou Kying Industrial Materials Co. Ltd склало порядку 0,37 ºС/Вт через значно нижчі величини теплопровідності ПІ діелектриків і більшу їхню товщину. При цьому в Моделі 1 температура в області джерела тепла потужністю 2 Вт досягала 104,1 та 134,4 ºС після 20 хвилин, переходячи в стійкий стан відповідно за сприятливих умов і за несприятливих умов експлуатації. У Моделі 2 температура в області джерела тепла потужністю 2 Вт досягала 116,1 та 146,3 ºС після 20 хвилин, переходячи в стійкий стан відповідно за сприятливих і за несприятливих умов експлуатації.

Таким чином, теоретичні дослідження теплових властивостей електронних модулів на комбінованих платах на алюмінієвих основах з розмірами 35 × 35 мм з поліімідними діелектриками на основі теплопровідної термозварювальної ПМФ плівки Kapton®120FMT616 компанії DuPont з теплопровідністю 0,46 Вт/(м·К) та на основі розробленого авторами композитного одностороннього лакофольгового мідь-поліімідного діелектрику з розрахунковою теплопровідністю ПІ плівки порядку 4,0 – 4,5 Вт/(м·К) показали можливість забезпечити робочі температури електронних модулів в області джерела тепла з підвищеною тепловою потужністю до 2 Вт в діапазоні порядку від 72,1 до 83,2 ºС для відносно сприятливих умов експлуатації модулів за природної конвекції при Та = 25 ºС, та в діапазоні порядку від 102,2 до 113,4 ºС для відносно несприятливих умов експлуатації при Та = 45 ºС (табл. 1).

2. Експериментальні дослідження теплових властивостей тестових структур якості електронних модулів на комбінованих платах з поліімідними діелектриками

Для експериментального фізичного моделювання та дослідження ефективності відводу тепла від напівпровідникових чипів з збільшеною потужністю в електронних модулях на комбінованих платах з ПІ ізоляцією, були виготовлені тестові структури якості (ТСЯ) електронних модулів на наступних відібраних типах комбінованих плат, в тому числі із застосуванням в якості діелектриків промислових тонких термозварювальних ПМФ плівок Kapton® 120FN 616 (ТСЯ 1), KYMIDE KYPIFEP 9198 (FHF) (ТСЯ 2) та Kapton® 120FMT 616 (ТСЯ 3), а також комбінованої плати на основі інноваційного одностороннього лакофольгового мідь-поліімідного діелектрику з високо теплопровідною ПІ композиційною плівкою з наповнювачем із суміші високодисперсних та ультрадисперсних порошків білого нітриду алюмінію (ТСЯ 4).

Комбіновані плати з ПІ ізоляцією із застосуванням промислових термозварювальних ПМФ плівок для ТСЯ 1, 2 та 3 виготовлялися шляхом приєднання до багатошарової ПМФ плівки з двох сторін мідної фольги зверху та алюмінієвої основи плати знизу за допомогою термообробки під тиском. Усереднені значення міцності на відшаровування мідної фольги від ПІ діелектрика у зразків, що виготовлялися, склали не менше ніж 2,5 Н/см.

Комбінована плата на основі одностороннього лакофольгового мідь-поліімідного діелектрику з високотеплопровідним композиційним шаром для ТСЯ 4 виготовлялася шляхом приклеювання безадгезивного мідь-поліімідного лакофольгового діелектрика до алюмінієвої основи плати за допомогою шару однокомпонентного вологозатвердженого теплопровідного полідиметилсилоксанового адгезивного матеріалу Kafuter K-5204K. Силіконовий адгезивний матеріал забезпечив хорошу адгезію до теплопровідного композиційного ПІ шару не менше 1,5 Н/см при зрушенні внахльост. В свою чергу усереднені значення міцності на відшаровування мідної фольги від теплопровідного композиційного ПІ шару у зразків, що виготовлялися, склали не менше 0,5 Н/см. Перевірка міцності на відшаровування мідної фольги від ПІ діелектриків у досліджуваних комбінованих платах проводилася відповідно до стандарту IPC-TM-650, метод тестування 2.4.9 [10].

В структурах ТСЯ для експериментальних досліджень теплових властивостей електронних модулів з відібраними типами комбінованих плат з розмірами 35 х 35 мм, в якості одиничних джерел тепла з потужністю 2 Вт, застосовувалися потужні світлодіоди серії SZ8-Y22-W0-C7-P компанії Seoul Semiconductor (Корея). Встановлення одиничних джерел тепла на мідні контактні площини верхніх шарів комбінованих плат проводили пайкою за допомогою паяльної пасти NC 293+ (Sn62/Pb36/Ag2) компанії AIM Solder (Канада) [11].

Експериментальні дослідження теплових властивостей відібраних варіантів ТСЯ електронних модулів проводилися у кімнатних умовах при природній конвекції та температурі навколишнього середовища T_a = 25°C, а також у не дуже прийнятних умовах експлуатації при температурі навколишнього середовища T_a = 45 °C в лабораторній електропечі СНОл-6.6.6/350 ГЦ-00.04. На рис. 3 представлено зовнішній вигляд ТСЯ для перевірки теплових властивостей електронних модулів на комбінованих платах з поліімідними діелектриками. На рис. 4 представлено графіки залежності температур нагріву комбінованих плат ТСЯ електронних модулів в області джерел тепла T_{hs} біля теплового контакту основ світлодіодів SZ8-Y22-W0-C7-P з платами та температур зворотної сторони алюмінієвої основи плат T_{Al} від часу до переходу їх у стійкий температурний стан.

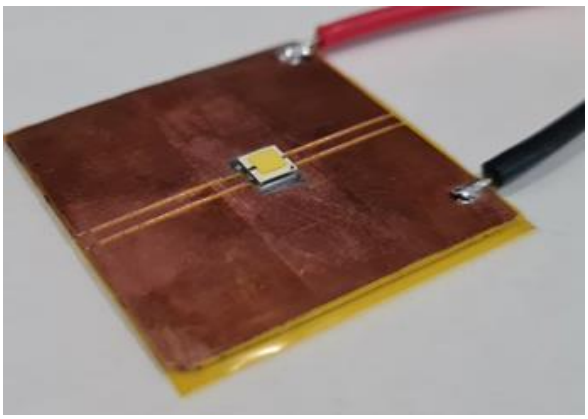


Рис. 3. ТСЯ для перевірки теплових властивостей на комбінованих платах з поліімідними діелектриками

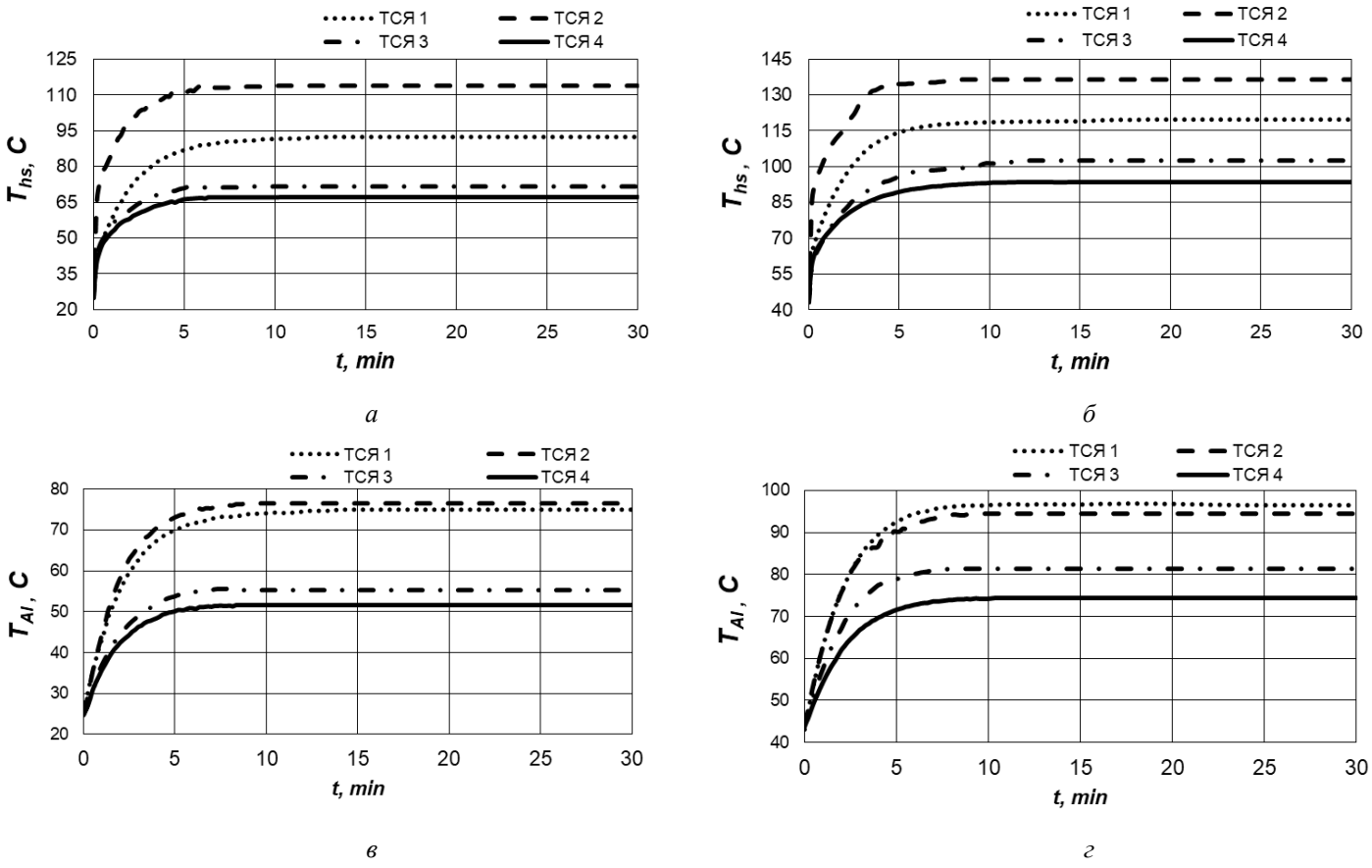


Рис. 4. Залежності температур нагріву комбінованих плат ТСЯ електронних модулів в області джерел тепла T_{hs}: *а* – температура довкілля T_a = 25°C; *б* – температура довкілля T_a = 45°C та температур нагріву зворотної сторони алюмінієвої основи плат T_{Al}: *г* – температура довкілля T_a = 25°C; *д* – температура довкілля T_a = 45°C від часу при переході у стабільний температурний стан

В табл. 2 представлено результати експериментальних досліджень температур нагріву плат ТСЯ електронних модулів в області паяння світлодіодів на плату та температур нагріву зворотної сторони алюмінієвих основ плат від часу при переході у стабільний температурний стан.

Таблиця 2

Результати експериментальних досліджень температур нагріву плат ТСЯ електронних модулів в області паяння світлодіодів на плату та температур нагріву зворотної сторони алюмінієвих основ плат від часу при переході у стабільний температурний стан при температурі довкілля *Ta = 25°C* та температурі довкілля *Ta = 45°C*

№ п/п	Варіанти ТСЯ	Температура на платі в області джерела тепла, T_{hs} °C		Температура на тильній стороні основи плати, T_{Al}, °C	
		Ta = 25°C	Ta = 45°C	Ta = 25°C	Ta = 45°C
1	ТСЯ 1	92,5	119,6	74,9	96,5
2	ТСЯ 2	113,9	136,6	76,6	94,4
3	ТСЯ 3	71,8	102,7	55,2	81,3
4	ТСЯ 4	67,2	93,5	51,6	74,4

3. Результати та їх обговорення

За отриманими експериментальними даними, які представлені на рис. 4 та в табл. 2, було підтверджено основні висновки теплового моделювання властивостей досліджених електронних модулів, яке було проведено за допомогою програмного комплексу COMSOL MULTIPHYSICS. Використання у конструкції комбінованих плат богатоошарових ПІ діелектриків з товщиною 30 мкм на основі термозварювальної ПМФ плівки Kapton®120FMT616 компанії DuPont з теплопровідністю 0,46 Вт/(м·К) дозволило суттєво покращити теплові властивості електронних модулів з збільшеною потужністю. На досліджених комбінованих платах з розмірами 35 х 35 мм в області джерела тепла площею ~ 4,29 мм^2 з потужністю 2 Вт за сприятливих умов експлуатації при температурі навколишнього середовища Ta = 25°C у ТСЯ було досягнуто температуру у стійкому стані порядку 71,8°C, а за несприятливих умов експлуатації при температурі навколишнього середовища Ta = 45°C в області пайки джерела тепла на плату температура у стійкому стані досягла 102,7°C. Застосування у конструкції комбінованих плат інноваційних односторонніх лакофольгових мідь-поліімідних діелектриків з високотеплопровідним композиційним ПІ шаром з товщиною 60 мкм з розрахунковою теплопровідністю порядку 4,0 – 4,5 Вт/(м•К) дало можливість забезпечити на платі за сприятливих умов експлуатації при температурі навколишнього середовища Ta = 25°C в області джерела тепла температуру у стійкому стані порядку 67,2°C та температуру порядку 93,5°C у стійкому стані за несприятливих умов експлуатації при температурі навколишнього середовища Ta = 45°C.

На досліджених комбінованих платах з використанням богатоошарових ПІ діелектриків на основі термозварювальної ПМФ плівки Kapton® 120FN 616 компанії DuPont товщиною 30 мкм з теплопровідністю 0,12 Вт/(м·К) температура в області джерела тепла потужністю 2 Вт досягала 92,5 та 119,6 °C після 20 хвилин, переходячи у стійкий стан відповідно за сприятливих умов та при несприятливих умовах експлуатації. На комбінованих платах з використанням богатоошарових ПІ діелектриків на основі термозварювальної багатошарової ПМФ плівки KYMIDE KYPIFEP 9198 (FHF) компанії Suzhou Kying Industrial Materials Co.Ltd (Китай) товщиною 50 мкм з мінімальною теплопровідністю 0,12 Вт/(м·К) температура в області джерела тепла потужністю 2 Вт досягала 113,9 та 136,6°C після 20 хвилин, переходячи у стійкий стан відповідно за сприятливих умов та при несприятливих умовах експлуатації. Таким чином, конструктивно-технологічні рішення комбінованих плат з використанням богатоошарових ПІ діелектриків на основі термозварювальної ПМФ плівки Kapton® 120FN 616м з фторполімерними двосторонніми покриттями товщиною 30 мкм з мінімальною теплопровідністю 0,12 Вт/(м·К) та плат на основі термозварювальної ПМФ плівки KYMIDE KYPIFEP 9198 (FHF) з фторполімерними двосторонніми покриттями товщиною 50 мкм з мінімальною теплопровідністю 0,12 Вт/(м·К) забезпечили за сприятливих умов експлуатації

робочі температури від 90 до 115 °С, а за несприятливих умов експлуатації лише від 120 до 135 °С. Такі конструкції комбінованих плат придатні до використання в електронних модулях з напівпровідниковими чипами лише з допустимими максимальними температурами від 125 до 145 °С.

Технічні рішення комбінованих плат на основі багатошарової теплопровідної термозварювальної ПМФ плівки Kapton®120FMT616 товщиною 30 мкм з фторполімерними двосторонніми покриттями з теплопровідністю 0,46 Вт/(м·К) та комбінованих плат на основі удосконалених односторонніх лакофольгових мідь-поліімідних діелектриків з високотеплопровідним композиційним ПІ шаром з товщиною 60 мкм з розрахунковою теплопровідністю до 4,0 – 4,5 Вт/(м•К), забезпечили за сприятливих умов експлуатації найкращі теплові характеристики електронних модулів з точки зору рекомендованих робочих температур < 80 °С для підтримки їх високої надійності роботи та підвищення строків експлуатації. При цьому удосконалена комбінована плата на основі інноваційного одностороннього лакофольгового мідь-поліімідного діелектрика з високотеплопровідним композиційним ПІ шаром за сприятливих умов експлуатації забезпечила рабочі температури < 70 °С. За несприятливих умов експлуатації конструкції плат забезпечили температури від 93,5 до 102,7ºС. Такі комбіновані плати придатні до використання в електронних модулях з напівпровідниковими чипами з допустимими максимальними температурами 110 °С та менше (табл. 2).

Комбіновані плати з ПІ ізоляцією із застосуванням промислових термозварювальних ПМФ плівок з двох сторонніми фтор-полімерними покриттями виготовлялися шляхом приєднання до багатошарової ПМФ плівки з двох сторін мідної фольги зверху та алюмінієвої основи плати знизу за допомогою термообробки при ~ 270 – 280°°С під тиском ~ 5 – 10 МПа. Виготовлення плат за методом термокомпресії є достатньо складним та енергозатратним, тому було розроблено нову конструкцію комбінованих плат та метод їх виготовлення на основі удосконалених високотеплопровідних односторонніх лакофольгових мідь-поліімідних діелектриків, які суттєво спростили та зменшили витрати на процес виготовлення комбінованих плат із забезпеченням їх високих теплових властивостей за рахунок приклеювання безадгезивних мідь-поліімідних лакофольгових діелектриків до алюмінієвих основ плат за допомогою тонкого шару однокомпонентного вологозатвердженого теплопровідного полідиметилсилоксанового адгезивного матеріалу з теплопровідністю ~1,6 Вт/(м·К).

Висновки

В роботі запропоновано конструкції та виконані теоретичні дослідження теплових моделей електронних модулів з підвищеною потужністю на основі комбінованих плат з використанням серійних термозварювальних поліімід-фторопластовими плівок, у тому числі з теплопровідністю від 0,12 до 0,46 Вт/м·К, а також на основі лакофольгових діелектриків з теплопровідністю композиційних ПІ шарів порядку 4,0 – 4,5 Вт/(м•К). Технічні рішення комбінованих плат на основі багатошарової теплопровідної термозварювальної ПМФ плівки Kapton®120FMT616 компанії DuPont товщиною 30 мкм з фторполімерними двосторонніми покриттями з теплопровідністю 0,46 Вт/(м·К) та комбінованих плат на основі удосконалених односторонніх лакофольгових мідь-поліімідних діелектриків з товщиною високотеплопровідних композиційних ПІ шарів до 60 мкм з теплопровідністю до 4,0 – 4,5 Вт/(м•К), забезпечують за сприятливих умов експлуатації при природній неутрудненій конвекції та температурі навколишнього середовища Та = 25ºС найкращі теплові характеристики електронних модулів з точки зору рекомендованих робочих температур < 70 – 80 °С для підтримки їх високої надійності роботи та підвищення строків експлуатації.

Список літератури:

1. Максимов А. Порівняльне дослідження теплопровідних властивостей матеріалів // Напівпровідникова світлотехніка. 2013. №4. С. 13–15.

2. Поліімідна плівка DuPont™ Kapton® HN, https://www.dupont.com/products/kapton-hn.html // офіційний сайт.

3. Комбіновані теплопровідні плати з діелектриками з поліміду / В.М. Борщов, О.М. Лістратенко, М.А. Проценко, І.Т. Тимчук, О.В. Кравченко, О.В. Суддя, І.В. Борщов, М.І. Сліпченко // Радіотехніка. 2023. Вип. 212. С. 11 –126.

4. Полімідно-фторопластова плівка DuPont™Kapton® 120FN616, https://www.dupont.com/products/kapton-fmt.html // офіційний сайт.

5. Полімідно-фторопластова плівка, що термозварюється, KYMIDE KYPIFEP 9198 (FHF) (Китай), https://www.kying.com // офіційний сайт.

6. Теплопровідна полімідно-фторопластова плівка DuPont™ Kapton® 120FMT616, https://www.dupont.com/products/kapton-fmt.html // офіційний сайт.

7. Нові підходи для створення ефективних комбінованих друкованих плат на теплопровідних основах з діелектриками з поліміду / В.М. Борщов, О.М. Лістратенко, М.І. Сліпченко, М.А. Проценко, І.Т. Тимчук, О.В. Кравченко, І.В. Борщов // Радіотехніка. 2023. Вип. 215. С. 60–68.

8. Structural modeling and calculation of thermal conductivity of polyimide composite materials / V.M. Borshchov, O.M. Listratenko, M.A. Protsenko, I.T. Tymchuk, O.V. Kravchenko, O.V. Syddia, I.V. Borshchov, M.I. Slipchenko // Radiotekhnika. 2022. №211. P. 133–142.

9. Клей герметик силіконовий теплопровідний Kafuter K-5204 (Китай), https: www.kafuter.cn/product-item-106.html // офіційний сайт.

10. Стандарт IPC-TM-650:2002. Test Methods Manual. Посібник із вибору методів контролю друкованих плат.

11. Паяльна паста марки NC293+ компанії AIM Solder, www.aimsolder.com // офіційний сайт.

Надійшла до редколегії 10.05.2024

Відомості про авторів:

Борщов Вячеслав Миколайович – д-р техн. наук, професор, ТОВ «Науково-виробниче підприємство «ЛТУ», перший заступник директора – головний конструктор; Україна; e-mail: viatcheslav.borshchov@cern.ch; ORCID: https://orcid.org/0000-0002-5579-8932

Лістратенко Олександр Михайлович – канд. техн. наук, ТОВ «Науково-виробниче підприємство «ЛТУ», провідний науковий співробітник; Україна; e-mail: sasha.listratenko.12@gmail.com; ORCID: https://orcid.org/0000-0001-7643-5295

Сліпченко Микола Іванович – д-р фіз.-мат. наук, професор, Інститут сцинтиляційних матеріалів НАНУ, провідний науковий співробітник; Україна; e-mail: naukovets.big@gmail.com; ORCID: https://orcid.org/0000-0002-4242-4800

Проценко Максим Анатолійович – канд. техн. наук, ТОВ «Науково-виробниче підприємство «ЛТУ», начальник відділення – заступник головного конструктора; Україна; e-mail: max.protsenko.1978@gmail.com; ORCID: https://orcid.org/0000-0001-9313-1701

Тимчук Ігор Трохимович – канд. техн. наук, ТОВ «Науково-виробниче підприємство «ЛТУ», головний технолог; Україна; e-mail: ihortymchuk78@gmail.com; ORCID: https://orcid.org/0000-0002-6436-7253

Кравченко Олександр Вікторович – ТОВ «Науково-виробниче підприємство «ЛТУ», заступник начальника відділу; Україна; e-mail: kravcenkoaleksandr671@gmail.com; ORCID: https://orcid.org/0000-0002-7145-4304

Борщов Ілля Вячеславович – ТОВ «Науково-виробниче підприємство «ЛТУ», інженер; Україна; e-mail: illia.borshchov1@nure.ua**; ORCID:** https://orcid.org/0000-0002-6598-6988.

УДК 621.383 **DOI:10.30837/rt.2024.2.217.14**

О.А. ЗАРУДНИЙ, канд. техн. наук

ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ ХАРАКТЕРИСТИК ВИПРОМІНЮВАЧА РЕЗОНАНСНОГО ЛІДАРУ З ОДНОПРОХОДОВИМ ПІДСИЛЮВАЧЕМ

Вступ

Резонансні лідари широко використовуються для досліджень верхньої атмосфери в інтересах геофізики, метеорології та екології навколишнього середовища. Особлива увага приділяється дослідженням вмісту атомів та іонів металів, що утворюються на висотах більше 80 км за рахунок абляції метеорної речовини [1 – 3].

Верхні шари атмосфери заповнені хвилями, які зароджуються в нижніх атмосфери від штормових систем і потоків повітря над гірською місцевістю, а потім поширюються вгору. Ці хвилі відіграють ключову роль у встановленні складу та температурних структур шляхом перемішування верхніх шарів атмосфери і збурюючими хімічними реакціями важливих реактивних речовин, таких як атомарний кисень і озон. З допомогою метеоритного натрію Na, що потрапляє в атмосферу шляхом випаровування космічного пилу на висоті від 78 до 110 км і який використовується як індикатор таких вторинних гравітаційних хвиль, проводяться дослідження цих хвиль резонансними лідарами [4 – 6]. Результати таких досліджень можуть також бути використані, наприклад, для виміру висотного розподілу вмісту аерозолів у верхній тропосфері та стратосфері [7].

Відомо, що передавач резонансного лідару, призначеного для дослідження домішок верхньої атмосфери, повинен володіти малою розбіжністю випромінювання, вузькою шириною спектральної лінії, можливістю плавної перебудови довжини хвилі випромінювання і великою енергією в імпульсі [8, 9]. Рідинні лазери на органічних барвниках з дисперсійним резонатором і ламповим накачуванням найкраще підходять для цієї мети, оскільки мають можливість перестроювання довжини хвилі випромінювання [8, 10].

З лідарного рівняння випливає [9, 10], що основними характеристиками передавача, які визначають ефективність резонансного лідара є енергія і смуга лінії випромінювання, безпосередньо пов'язана з ефективним перетином розсіювання речовини [11]. Однак між цими двома характеристиками лазерного генератора існує параметрична зв'язок, тобто, прагнення до звуження смуги генерації призводить до неминучого зниження енергії випромінювання через додаткові втрати, що вносяться в резонатор елементами селекції. У роботах [12, 13] показано, що при оптимальному виборі характеристик дисперсійного резонатора можливе досягнення максимальної ефективності використання енергії випромінювання передавача при взаємодії з атомами атмосферної домішки, що досліджується.

Для подальшого підвищення енергетичних характеристик передавача лідара у схемі випромінювача можуть бути використані оптичні підсилювачі. Оскільки для резонансного зондування атмосфери принципове значення має спектральна чистота випромінювання переважно використання підсилювачів біжучої хвилі [11, 14, 15]. У цьому випадку спектральні та просторові характеристики вихідного пучка випромінювача змінюються незначно, оскільки в однопроходових підсилювачах, на відміну від багатопроходових, немає накопичення аберацій. Крім того, надійність однопроходових підсилювачів забезпечується відсутністю дзеркальних покриттів. Теоретичний аналіз показав високу ефективність їх застосування у схемі генератор-підсилювач [11].

Мета роботи – експериментальна перевірка результатів теоретичного аналізу спектрально-енергетичних параметрів випромінювача з ламповим накачуванням, побудованого за схемою генератор-підсилювач, отриманих раніше [11].

Методика експерименту

В експериментах використовувалася конструкція лазера [10], з плоско-паралельним резонатором. Як генераційне середовище використовувався спиртовий розчин барвника родамін 6Ж. Інтервал зміни довжин активних елементів генератора та підсилювача вибирався рівним 12 см, що відповідало розрядному проміжку ламп ІСП-5000, які використовувалися в експериментальних дослідженнях.

Як внутрішньорезонаторні селектори в експериментах використовувалися юстируємі інтерферометри Фабрі – Перо з пластинами середньої оптичної якості (якість обробки ~ $\lambda/50$ на діаметрі 40 мм) [10]. Відхилення в товщині зазору по діаметру пучка випромінювання визначалося в основному якістю юстування дзеркал. При цьому налаштування інтерферометрів здійснювалося на стенді вручну таким чином, щоб величина кута нахилу при налаштуванні на резонансну лінію знаходилася в межах 5 – 10 мрад. Доцільність використання юстованих інтерферометрів замість цілісних еталонів очевидно випливає з порівняння внесених ними неселективних втрат. Попередні розрахунки показують [12], що втрати, внесені юстованими інтерферометрами при малих кутах нахилу виявляються значно меншими, ніж цілісних еталонів вищої якості, для яких ці кути, у загальному випадку, значно більше. Виняток становить варіант, в якому як останній ступінь селекції застосовуються еталони з базою близько декількох міліметрів. У цьому випадку граничні кути нахилу для одного порядку інтерференції невеликі та використання цілісного еталону виправдане зручністю експлуатації та стабільністю характеристик.

Як основний селектор у дослідженнях використовувалася конструкція вимірювального інтерферометра ІТ-51, з інваровими прокладками, в якому пластини з кварцового скла мали діелектричні просвітлювальні і відбиваючі покриття. Перевищення бази резонатора над довжиною активного елемента генератора дорівнювало 85 см. Для протидії паразитної модуляції спектральної лінії вихідне дзеркало і найближчий інтерферометр були віддалені від вікон кювети на відстань не менше 25 см. З цією ж метою вікна кювети з активною рідиною були просвітлені та нахилені один до одного та до осі резонатора на кути не менше 0,5 град.

Спектр випромінювання розраховувався за допомогою машинної обробки реєстрованих ПЗЗ матрицею інтерферограм, отриманих за допомогою вимірювального інтерферометра ІСП-51 і об'єктива з фокусною відстанню 800 мм. Енергія випромінювання вимірювалася за допомогою вимірювача калориметричного типу ІКТ-1М.

Оптична схема лазера з дисперсійним резонатором, що використовувалася в експериментах представлена на рис. 1.

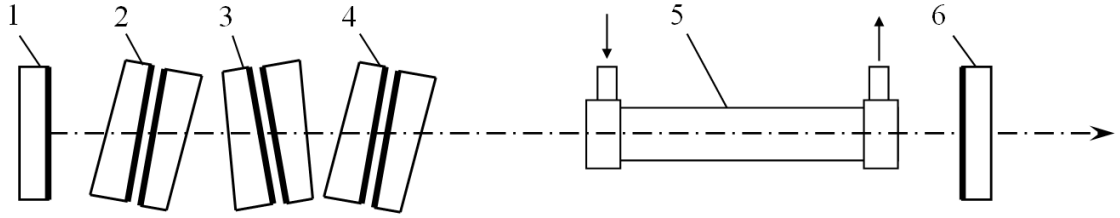


Рис. 1. Схема розташування оптичних елементів генератора

На рис. 1 позначено: 1 – «глухе» дзеркало; 2, 3, 4 – інтерферометри-селектори; 5 – кювета з барвником; 6 – вихідне дзеркало резонатора.

Тут доречно зробити пояснення з приводу встановлення селекторів у резонаторі, оскільки можливі два варіанти їх розміщення: між «глухим» дзеркалом і кюветою і між кюветою і вихідним дзеркалом. При малих втратах в резонаторі це немає значення, тому розташування

селектору у переважній більшості випадків не має особливого значення. Однак ситуація різко змінюється для великих втрат, що як раз є характерним для даного випадку.

При значних втратах на виведення випромінювання (малій величині коефіцієнта відбиття вихідного дзеркала) і з урахуванням неселективних втрат, що вносяться внутрішньорезонаторними інтерферометрами не можна нехтувати зміною інтенсивності випромінювання по довжині резонатора, тому з точки зору ККД стаціонарної генерації зовсім не байдуже, яким чином розташований селектор всередині резонатора.

Відомо, що потужність квазістаціонарної генерації будь-якого лазера визначається виразом [13]

$$P_г = {\nu_г}/{\nu_н} \cdot p_0 \cdot p_1 \cdot p_3 \cdot \left(P_н - P_{нп}\right),$$

де $\nu_г, \nu_н$ – відповідно частоти генерації та накачування; p_0 – ККД системи накачування; p_1 – частина енергії накачування, що збуджує молекули на робочий рівень; p_2 – відношення імовірностей спонтанних і вимушених переходів; p_3 – коефіцієнт, що характеризує співвідношення корисних і шкідливих втрат в резонаторі; $P_н$ – потужність накачування; $P_{нп}$ – порогова потужність накачування.

Розглянемо спочатку варіант розміщення селектора так, як показано на рис. 1. Якщо посилення кювети з барвником за один прохід позначити K, інтенсивність генерації $I_г$, коефіцієнт відбиття вихідного дзеркала R, а коефіцієнт пропускання селектора T_c, то можна показати, що втрати потужності в селекторі через його неповне пропускання будуть рівні

$$\left[I_г \cdot R/(1-R)\right] \cdot K \cdot \left(1 - T_c^2\right).$$

Врахуємо також, що у стаціонарному режимі має виконуватися умова

$$K^2 \cdot T_c^2 \cdot R = 1.$$

Отже, згідно з визначенням p_3 отримаємо вираз

$$p_3 = \frac{1}{1 + \dfrac{\sqrt{R}}{1-R} \cdot \left(\mathrm{T}_c^{-1} - \mathrm{T}_c\right)}.$$

Аналогічно для варіанта розміщення селектора між кюветою та вихідним дзеркалом, отримаємо

$$p_3 = \frac{1}{1 + \dfrac{\sqrt{R}}{1-R} \cdot \left(\mathrm{T}_c^{-1} - \mathrm{T}_c\right)}.$$

Розрахунки, проведені з використанням отриманих виразів, показують, що завжди енергетично вигідніше встановити селектор між «глухим» дзеркалом і кюветою з активною рідіною. Це зумовлено зміною сумарної інтенсивності зустрічних хвиль по довжині резонатора. При цьому чим менша сума інтенсивностей, тим, зрозуміло, менше втрати, які вносять селектор. Тому в експериментах використовувався варіант розміщення селекторів, показаний на рис. 1 і в наступному порядку (рахуючи від «глухого» дзеркала): інтерферометр з базою 2, 0,3 і 8 мкм.

Для генератора в експериментах використовувалася конструкція лазерної головки, яка забезпечувала можливість зміни довжини активного елементу із збереженням постійної щільності енергії накачування. Довжина підсилювача варіювалася з інтервалом 12 см у діапазоні 24 – 60 см [11].

Як основний критерій оцінки ефективності системи генератор-підсилювач використовувалася величина ефективної випромінюваної енергії, яка визначається виразом [12]

$$E_{еф} = E_0 \cdot \sigma_{еф} / \sigma_{\max},$$

де E_0 – енергія випромінювання передавача; $\sigma_{еф}$ – ефективний перетин розсіювання на атомах домішки; $\sigma_{\max}$ – перетин поглинання (розсіювання) у максимумі лінії.

Очевидно, що $E_{еф}$ характеризуватиме енергію випромінювання лазера, що потрапляє в спектр поглинання домішки, яка досліджується. Величина ефективної випромінюваної енергії була обрана в якості основного критерію оцінки ефективності роботи передавача резонансного лідару, оскільки згідно з рівнянням лазерної локації [8, 9] визначає рівень прийнятого сигналу.

Експериментальні дослідження та обговорення результатів

На рис. 2 представлено результати вимірювань для системи з чотириламового генератора з дисперсійним резонатором, характеристики якого представлені вище і підсилювача. Протяжність підсилювача в експерименті була незмінною і дорівнювала 600 мм. Концентрація барвника в експериментах вибиралася з умови $\sigma_{01}^{max} rn \approx 5$, що давало можливість використання як для генератора, так і для підсилювача загальної системи прокачування активної рідини. За допомогою каліброваних світлодільників та світлофільтрів енергія на вході підсилювача змінювалася від 1 до 0,3 Дж. Тут же наведено результати теоретичних розрахунків [11] за вихідних даних, що відповідають експерименту (пунктирна крива). У розрахунках враховувалося зниження ККД накачування на 10 %, пов'язане із збільшенням навантаження на розрядник. З рис. 2 видно, що результати розрахунків та експериментальні результати є близькими.

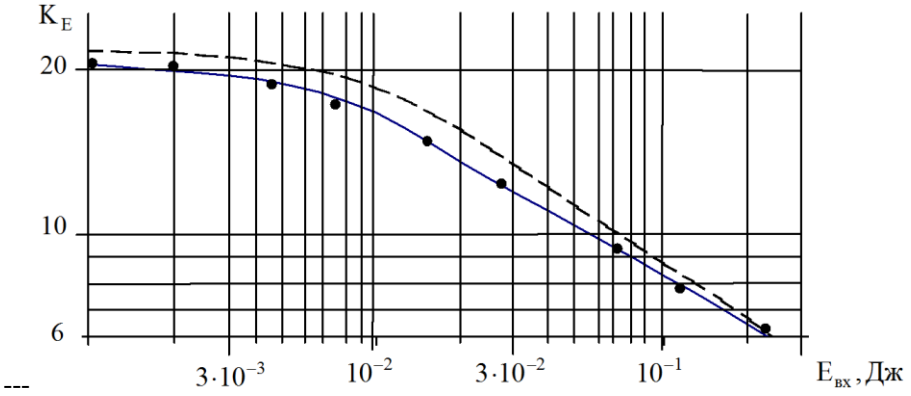


Рис. 2. Експериментальна залежність коефіцієнту підсилення від вхідної енергії

Найкраще узгодження спостерігається в області високих енергій, коли роль посиленої люмінесценції в балансі потужностей невелика. Дещо гірше узгодження в області малих вхідних енергій можна пояснити невідповідністю моделі рівномірного розподілу інтенсивності посиленого шуму реальної ситуації, що має місце в підсилювачі біжучої хвилі. Інтенсивність посиленого шуму на краях кювети значно перевищує інтенсивність у середині [12], тому хід розрахункової залежності коефіцієнта посилення в області малих енергій дещо відрізняється від експериментальної кривої у бік менших значень.

На рис. 3 представлено експериментальні результати досліджень ефективної енергії випромінювання системи генератор-підсилювач $E_{еф}$ (суцільна лінія) при зміні довжини генератора $\ell_г$ при збереженні сумарної протяжності генератора і підсилювача, що дорівнює 84 см. Тут же для порівняння представлена залежність енергії випромінювання E в широкосмуговому варіанті (пунктир) за тих самих умов експерименту.

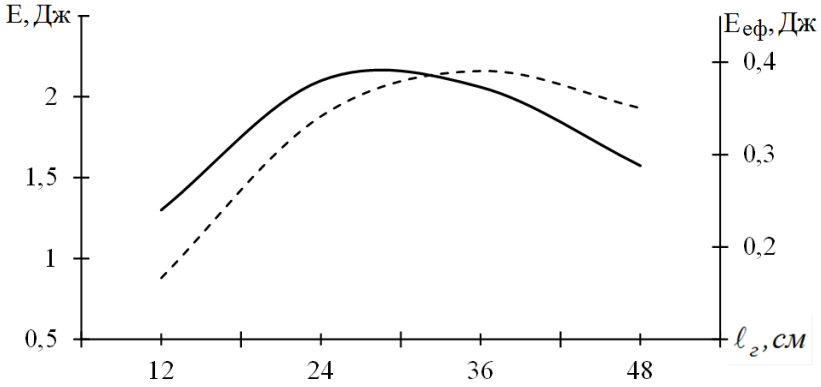


Рис. 3. Залежність енергетичних характеристик випромінювача від довжини генератору

Коефіцієнт відбиття вихідного дзеркала генератора у кожному випадку підбирався з умови отримання максимальної енергії випромінювання генератора. Порівняння представлених на рис. 3 кривих показує відмінність оптимальних співвідношень довжин генератора і підсилювача щодо різних критеріїв. Якщо ставиться завдання отримання максимальної енергії випромінювача з широкосмуговим варіантом генератора, оптимальним є приблизно однакові довжини генератора і підсилювача. Якщо система оптимізується за $E_{еф}$, то оптимальна довжина генератора змінюється у бік менших значень. Це пов'язано, перш за все, зі зміною спектральної ширини випромінювання при зміні протяжності генератора. Зіставлення теоретичних [11, 12] та експериментальних залежностей показує їх згоду.

На рис. 4 представлено результати вимірювань кута розходження випромінювання θ системи генератор-підсилювач залежно від довжини підсилювача та при фіксованої довжини генератора 24 см, що є близькою к оптимальній.

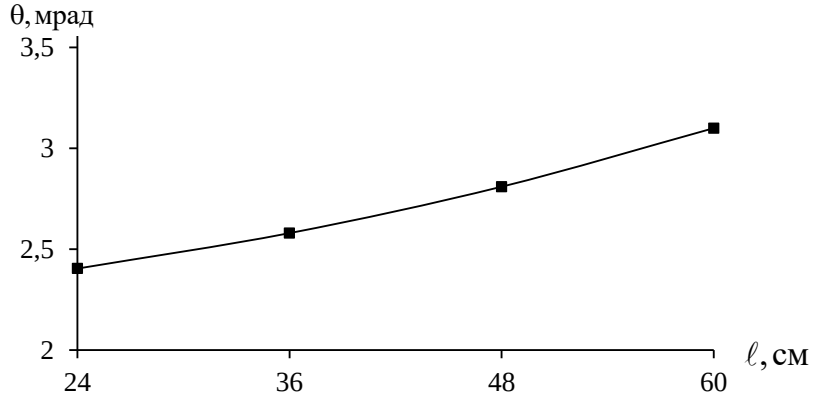


Рис. 4. Залежність кута розходження випромінювання від довжини підсилювача

З рис. 4 видно, що кут розходження випромінювання монотонно зростає від вихідних 2,3 мрад генератора до приблизно 3 мрад системи генератор-підсилювач при максимальній протяжності підсилювача. Порівняно незначне збільшення розбіжності випромінювання обумовлено наведеними термооптичними спотвореннями в активному елементі підсилювача при ламповому накачуванні і може бути компенсовано застосуванням коліматора в оптичній схемі передавача лідару.

Висновки

Для лідарних досліджень верхньої атмосфери резонансним способом необхідний випромінювач лідару значної потужності і вузької смуги спектру випромінювання. Використання одного лише генератору в схемі передавача обмежує можливості збільшення спектральної яскравості випромінювання. Використання в схемі випромінювача оптичного підсилювача дозволяє при відносно невеликій потужності генератора досягти значного покращення його

енергетичних характеристик. Результати експериментальних досліджень однопроходового підсилювача дозволяють зробити висновок про суттєве підвищення енергетичного потенціалу резонансного лідару із зберіганням спектральної чистоти випромінювання. При цьому отримано незначне зростання кута розходження випромінювання. Збільшення протяжності накачуваної області однопроходового підсилювача має обмеження переважно за рахунок підвищення негативного впливу посиленої широкосмугової люмінесценції. Експериментальні дослідження системи генератор-підсилювач підтверджують висновки теоретичного аналізу і свідчать про те, що при фіксованій сумарній протяжності накачуваної області існує оптимальне співвідношення довжин генератора та підсилювача, при якій досягається максимальна величина ефективної енергії випромінювання передавача.

Список літератури:

1. Chu X., Nishimura Y., Xu Z., Yu Z., Plane J. M. C., Gardner C. S., & Ogawa Y. (2020). First simultaneous lidar observations of thermosphere-ionosphere Fe and Na (TIFe and TINa) layers at McMurdo (77.84°S, 166.67°E), Antarctica with concurrent measurements of aurora activity, enhanced ionization layers, and converging electric field // Geophysical Research Letters, 47, e2020GL090181. doi: 10.1029/2020GL090181.
2. Chu X., Chen Y., Cullens C. Y., Yu Z., Xu Z., Zhang S.-R. et al. (2021). Mid-latitude thermosphere-ionosphere Na (TINa) layers observed with high-sensitivity Na Doppler lidar over Boulder (40.13°N, 105.24°W) // Geophysical Research Letters, 48, e2021GL093729. вої: 10.1029/2021GL093729
3. Swenson G. R., Salinas C. C. J. H., Vargas F., Zhu Y., Kaufmann M., Jones M. Jr., et al. (2019). Determination of global mean eddy diffusive transport in the mesosphere and lower thermosphere from atomic oxygen and carbon dioxide climatologies // Journal of Geophysical Research: Atmospheres, 124, 13,519–13,533. doi:10.1029/2019JD031329.
4. Chu X., Gardner C. S., Li X., & Lin C. Y.-T. (2022). Vertical transport of sensible heat and meteoric Na by the complete temporal spectrum of gravity waves in the MLT above McMurdo (77.84°S, 166.67°E), Antarctica // Journal of Geophysical Research: Atmospheres, 127, e2021JD035728. doi: 10.1029/2021JD035728.
5. J. Wu, W. Feng, X. Xue, D. R. Marsh, J. M. C. Plane, X. Dou.The 27-Day Solar Rotational Cycle Response in the Mesospheric Metal Layers at Low Latitudes, Geophysical Research Letters, 10.1029/2019GL083888, **46**, 13, (7199-7206), (2019).
6. Kylee Branning, Mark Conde, Miguel Larsen, Riley Troyer, Resolving Vertical Variations of Horizontal Neutral Winds in Earth's High Latitude Space-Atmosphere Interaction Region (SAIR) // Journal of Geophysical Research: Space Physics, 10.1029/2021JA029805, **127**, 5, (2022).
7. Mikhalev A.V., Tashchilin M.A. and Sakerin S.M. Effect of Atmospheric Aerosol on Ground-Based Airglow Observations // Atmospheric and Oceanic Optics, 2019, V.32. No.04. pp.410–415.
8. X. Chu and G. Papen. Resonance fluorescence lidar for measurements of the middle and upper atmosphere // Laser Remote Sensing, T. Fujii, and T. Fukuchi, Eds., pp. 179–432, CRC Press (2005).
9. Norman Hodgson, Horst Weber. Laser Resonators and Beam Propagation: Fundamentals, Advanced Concepts and Applications. 2nd Edition. Springer, 2005.
10. Зарудный А.А., Плетенев В.Г., Верхоробин А.Л. Лазер повышенной спектральной яркости для исследования атмосферы // Радиотехника. 1998. Вып.102. С.170–175.
11. Zarudnyi A.A., Tsopa A.I. Power Characteristics of the Lidar Transmitter Assembled in Generator-Amplifier Circuit-Design // Telecommunication and Radio Engineering. 2019. Vol. 78(1). P. 31–37.
12. Спектральные характеристики передатчика резонансного лидара на основе лазера на красителях с дисперсионным резонатором / В.Л. Басецкий, А.А. Зарудный // Радиотехника. 2012. Вып. 169. С. 359–363.
13. Weber M. J. Handbook of Lasers // Laser & Optical Science & Technology. CRC Press, 2019.
14. Tunable polymer dye laser pumped by two 513 nm diodes / O. A. Burdukova [et al.] // Laser Physics Letters. 2020. Vol. 17, no. 2. P. 795–801.
15. Tunable dye laser amplifier chain for laser isotope separation / I. S. Grigoriev [et al.] // Quantum Electronics. 2004. Vol. 34, N.5. P. 447–450.

Надійшла до редколегії 15.05.2024

Відомості про автора:

Зарудний Олександр Андрійович – канд. техн. наук, Харківський національний університет радіоелектроніки, доцент кафедри радіотехнологій інформаційно-комунікаційних систем; Україна, e-mail: oleksandr.zarudnyi@nure.ua ; ORCID: https://orcid.org/0000-0002-1612-0256

УДК 621.396.96 **DOI:10.30837/rt.2024.2.217.15**

О.В. КАРТАШОВ, І.Є. КОНДРАШОВ

МЕТОД АДАПТАЦІЇ СИСТЕМ РАДІОАКУСТИЧНОГО ЗОНДУВАННЯ АТМОСФЕРИ

Вступ

Системи радіоакустичного зондування атмосфери (РАЗ) є ефективним та перспективним засобом отримання інформації про стан та динамічні процеси, що відбуваються в нижніх шарах атмосфери. Системи РАЗ дозволяють здійснювати вимірювання вертикальних профілів таких характеристик атмосфери як температура, швидкість та напрям вітру, параметри турбулентності, вологість повітря [1 – 4].

Теорія і практика систем РАЗ розвиваються протягом кількох десятиліть, починаючи з 1961 р. [5 – 7], проте до цього часу не вдалося подолати низку обмежень і недоліків, які суттєво обмежують можливості застосування систем РАЗ на практиці при вирішенні актуальних прикладних завдань – метеорологічне забезпечення зльоту та посадки літальних апаратів, прогнозування поширення радіо- та акустичних хвиль в атмосфері, прогнозування умов, що призводять до формування екологічно небезпечних ситуацій [3, 8, 9].

Основними серед існуючих обмежень систем РАЗ є порушення умов Брегга по трасі зондування та вітровий знос плями розсіяного радіосигналу внаслідок переміщення акустичного хвильового пакета під дією вітру. В даний час відомий ряд алгоритмів, спрямованих на усунення або компенсацію впливу порушення умови Брегга на точність та оперативність радіоакустичного зондування. Однак ці алгоритми не дозволяють суттєво покращити зазначені характеристики систем РАЗ, оскільки вони не враховують ряд особливостей розсіювання радіохвиль на неоднорідностях, створюваних акустичним хвильовим пакетом, та отримані евристичним шляхом [10 – 13].

У статті синтезується алгоритм частотної адаптації систем РАЗ до метеорологічної обстановки, що змінюється, на основі адекватної моделі радіоакустичного інформаційного каналу і з використанням основних досягнень теорії оптимального управління.

Відомі алгоритми частотної адаптації систем РАЗ

Принцип дії систем РАЗ заснований на розсіюванні електромагнітних хвиль на збуреннях середовища, створюваних імпульсними звуковими коливаннями, що випромінюються з поверхні землі у вертикальному напрямку.

Для того щоб отримати достатній для реєстрації та обробки рівень розсіяного радіосигналу, необхідно забезпечити виконання умови Брегга [1, 3], за якої електромагнітні хвилі, розсіяні різними частинами акустичного пакета, складаються з урахуванням їх фаз когерентно

$$\lambda_e = 2\lambda_s \sin\theta, \tag{1}$$

де λ_s – довжина звукової хвилі; λ_e – довжина електромагнітної хвилі; θ – кут розсіювання електромагнітних хвиль.

Довжина електромагнітної хвилі, що випромінюється, практично не залежить від значень параметрів атмосфери, а довжина хвилі акустичного випромінювання істотно змінюється внаслідок змін температури повітря з висотою. У зв'язку з цим виникає необхідність підстроювання частот зондувальних сигналів з метою виконання умови Брегга по трасі зондування. При цьому в принципі можна змінювати адаптивно частоту як акустичного, так і електромагнітного зондувальних коливань.

Технічно простіше реалізувати зміни частоти звукового сигналу. З початку розвитку методу РАЗ робилися досліди, спрямовані на вимірювання профілів температури атмосфери «за допомогою однієї звукової посилки», коли частота акустичного сигналу підбирається для метеорологічних умов, які існують у середній точці висотного профілю.

На рис. 1 наведено залежності дисперсії температури атмосфери від висоти, отримані експериментальним шляхом при використанні двох методик зондування [14, 15]. Графіки, що відповідають методиці «точка», показані суцільними лініями, а графік, отриманий з використанням методики «траса», – пунктирною лінією. В експерименті застосовувалася система РАЗ, що використовує такі частоти зондувальних сигналів: частота звукового сигналу – 5 см, частота радіосигналу – 10 см.

Методика зондування «траса» передбачає виконання умови Брегга спочатку на деякій нижній висоті, виконання заданої кількості зондувань атмосфери для цієї висоти, вимірювання зсувів доплерівської частоти з подальшим усередненням отриманих результатів вимірювання температури на обраній висоті. Далі описана сукупність дій послідовно виконується для наступних висот профілю. Відстань по висоті між сусідніми точками профілю зазвичай вибирається рівною просторової протяжності акустичного зондувального імпульсу.

Методика «точка» передбачає такий вибір частоти акустичного сигналу, при якій оптимальне співвідношення довжин хвиль акустичного та електромагнітного сигналів забезпечується на деякій середній висоті профілю. При цьому частота звукового сигналу підлаштовується під метеоумови, що спостерігаються на цій висоті, експериментально, по максимуму відбитого радіосигналу [14].

Далі здійснюється випромінювання пакета звукових хвиль та послідовна реєстрація значень доплерівських зсувів частоти через рівні інтервали часу, що відповідають переміщенню звукового пакета на величину його просторової протяжності, у міру поширення пакета звукових хвиль трасою зондування [15]. Зондування здійснювалося протягом часу, який зазвичай застосовується при усередненні отриманих «миттєвих» результатів вимірювань метеопараметрів (2, 3, 5, 10 хв.).

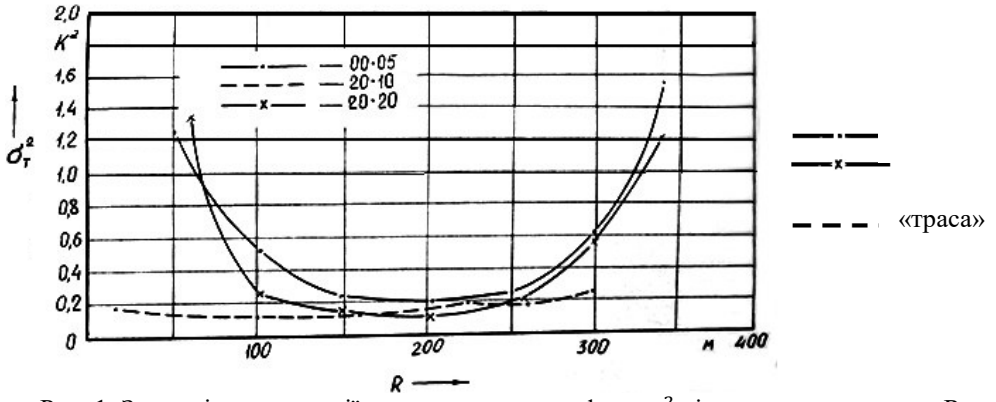


Рис. 1. Залежність дисперсії температури атмосфери σ_T^2 від висоти зондування R ри використанні методик зондування «точка» і «траса»

З даних рис. 1 видно, що дисперсія профілів температури, отриманих методом «траса», знаходиться приблизно на однаковому рівні, з деякою тенденцією до зростання, що спостерігається на найбільш висотній ділянці профілю. Що може пояснюватися зменшенням значень відношення сигнал шум для відбитого радіосигналу, що отримується з цих висот.

Дисперсія профілів температури, отриманих з використанням методики «точка», має мінімальні значення в районі висоти 200 м, для якої виконувалася настройка умови Брегга, що забезпечує отримання максимальної амплітуди радіосигналу, що відбивається від звукової посилки.

У міру відходу від цієї точки у бік більших і менших висот, спостерігається збільшення значення дисперсії температури, причому для великих висот абсолютні значення дисперсії дещо більше, що також пояснюється меншими значеннями співвідношень сигнал-шум для цих висот.

Забезпечення налаштування за умови Брегга в кожній з точок висотного профілю (які іноді називають «майданчиками») шляхом зміни частоти звукового сигналу зондуючого

вручну вимагає значної кількості часу, яка в залежності від наявних метеорологічних умов і від кількості висотних точок профілю може становити 1-2 години.

Такий алгоритм виконання вимірювань суттєво обмежує такий показник систем РАЗ, як оперативність отримання профілів метеоінформації. У той самий час даний показник є одним із основних потенційних переваг методу в порівнянні з наявними засобами вимірів. Крім того, зазначений час вимірів можна порівняти з часом квазістаціонарності процесів в атмосфері. Перевищення часу квазістаціонароності при виконанні вимірювань методом усереднення значень метеопараметра призводить до суттєвого погіршення результуючої похибки вимірювань. Саме з цієї причини вже на початковому етапі розвитку методу РАЗ стали робити спроби отримати профілі температури «по одній звуковій посилці» з підстроюванням частоти акустичного сигналу в середньому по трасі.

У літературі описані також алгоритми частотної адаптації систем РАЗ до метеообстановки, що змінюється, в яких забезпечення умови Брегга досягається шляхом підстроювання частоти зондуючого радіосигналу в міру просування зондуючого акустичного хвильового пакета в атмосфері. Це стає можливим через суттєву відмінність швидкостей використовуваних зондувальних коливань – акустичного та електромагнітного.

У роботі [16] описано використання методу фазового автопідстроювання частоти (ФАПЧ) радіосигналу при реалізації алгоритму частотної адаптації систем РАЗ. Використання автопідстроювання частоти радіосигналу системи заснована на тому факті, що при виконанні умови Брегга доплерівський зсув частоти радіосигналу точно співпадає зі значенням несучої частоти звукового зондувального сигналу f_s.

У даному методі, у міру поширення звукового імпульсу в напрямку зондування, здійснюється вимірювання параметра Δ – поточного значення різниці між значеннями несучої частоти розсіяного радіосигналу f_p і номінальної частоти задаючого генератора радіосигналу. Далі здійснюється перетворення значення параметра Δ в керуючу напругу $U_{\text{упр}}$, яка використовується для управління частотою задаючого генератора радіосигналу. Це дозволяє забезпечити виконання умови Брегга по всій трасі зондування.

Інформація про швидкість звуку в атмосфері в даному методі зондування буде міститися не у значенні частоти розсіяного сигналу $f_{\text{р}}$, як це зазвичай буває, а зміні несучої частоти радіосигналу. Значення параметра $f_{\text{р}}$ має залишатися незмінним.

Схема ФАПЧ включала в себе компаратор, фазовий детектор і пропорційно-інтегруючий фільтр. Поточні вимірювання значень частоти генератора електромагнітного випромінювання та зсувів частоти розсіяного сигналу вимірювалися в системі методом «рахунку нулів», далі отримані значення вводилися в комп'ютер, в якому обчислювалися значення швидкості звуку в атмосфері, а також значення температури середовища по трасі зондування.

У розглянутій схемі зондування за наявності значного початкового розстроювання частот зондувальних сигналів має місце зрив стеження схеми автопідстроювання за значенням f_s . Внаслідок цього потрібно здійснювати попередній вибір частоти акустичного сигналу з метою виконання умови Брегга в будь якій точці траси зондування. У разі сильного поривчастого вітру мають місце глибокі завмирання амплітуди розсіяного радіосигналу на окремих ділянках траси, і стійкість роботи системи радіоакустичного зондування у такому разі також порушується.

Відзначено [16], що за наявності зривів у роботі системи автопідстроювання частоти отримані значення температури атмосфери на кілька градусів перевищують значення температури у сусідніх точках профілю, тобто, по суті, мають місце аномальні похибки результатів вимірювань.

Таким чином, розглянуті алгоритми частотної адаптації систем РАЗ не забезпечують практично необхідної точності та оперативності вимірювань. Це пояснюється насамперед використанням невірних уявлень про процес розсіювання радіохвиль на протяжній радіолокаційної цілі у вигляді акустичного хвильового пакета. Крім того, алгоритми адаптації

створені евристичним шляхом, без використання знань про процеси управління об'єктами та спостереженнями, що отримані у відповідних галузях науки та техніки [2, 17 – 20].

У статті розглядається синтез алгоритму частотної адаптації систем РАЗ до метеорологічної обстановки, що змінюється, на основі адекватної моделі радіоакустичного інформаційного каналу і з використанням основних досягнень теорії стохастичного оптимального управління.

Математична модель радіоакустичного інформаційного каналу

Значна частина завдань щодо побудови теорії радіоелектронних систем у різних галузях пов'язана з побудовою математичної моделі інформаційного каналу, що описує механізм поширення та розсіювання хвиль у існуючих системах [21 – 23].

У сучасній теорії вимірювальних радіосистем зондований об'єкт, що досліджується, розглядається як елемент інформаційного каналу і представляється деяким детермінованим або стохастичним математичним оператором розсіювання, який описує зміни просторово-часового сигналу при його проходженні по каналу.

За наявності адекватного, конструктивного математичного оператора інформаційного каналу стає можливою побудова теорії радіосистем відповідного призначення (а також теорії зондувальних сигналів), суттєво спрощується вирішення завдань вибору видів сигналів при проєктуванні станцій, спрощується завдання оцінювання параметрів, керування системою [24 – 26].

В даний час при аналізі, синтезі та проєктуванні комплексів РАЗ найчастіше використовуються математичні уявлення, засновані на строгому вирішенні радіофізичного завдання на основі рівнянь Максвелла, внаслідок чого вони є досить складними та громіздкими [3]. Для фахівців у галузі технічних засобів РАЗ при розгляді обговорюваних у роботі завдань доцільно використовувати більш простіший і фізично наочний модельний підхід, заснований на математичному апараті, що використовується в теорії систем, що дозволяє відобразити характерні особливості процесу розсіювання радіохвилі на звуку та розсіяного інформаційного сигналу.

В роботі розглядається математична модель інформаційного локаційного каналу як взаємної кореляційної просторової функції акустичного і радіо сигналів за їх просторовим поданням [11, 12]

$$E_1(r) = \mathrm{A}\int_{-\infty}^{\infty} E(2r' - r)S^*(r')e^{jqr'}dr', \tag{2}$$

де E – комплексна обвідна зондувального радіосигналу; S – комплексна обвідна зондувального акустичного сигналу; $q = 2k_e - k_s$ – параметр розстроювання умови Брегга; k_e, k_s – хвильові числа відповідно електромагнитної та акустичної та хвиль; r – взаємне зміщення зондуючих атмосферу сигналів вздовж просторової координати r'; A – амплітудний множник пропорційності.

Застосувавши до правої частини виразу (2) теорему Парсеваля, отримаємо рівняння, де розсіяний сигнал визначається через просторові спектри відповідних комплексних обвідних коливань, що випромінюються системою

$$E_1(r) = \frac{K}{4\pi}\int_{-\infty}^{\infty} S_E\left(\frac{k}{2}\right)S_S^*(k+q)e^{-j\frac{r}{2}k}dk, \tag{3}$$

де $\int_{-\infty}^{+\infty} S(r')e^{-jqr'}e^{-jkr'}dr' = S_S(k+q)$,

$\int_{-\infty}^{\infty} E(2r' - r)e^{-jkr'}dr' = e^{-j\frac{r}{2}k}\int_{-\infty}^{\infty} E(2r')e^{-jkr'}dr' = \frac{1}{2}e^{-j\frac{r}{2}k}S_E\left(\frac{k}{2}\right)$,

$k = \frac{2\pi}{r'}$ – просторова частота.

У ряді випадків більш зручнішим є наступний вираз для розсіяного сигналу:

$$E_1(r) = \frac{K}{4\pi}\int_{-\infty}^{\infty} S_E(k)S_S^*(2k+q)e^{-jrk}dk, \tag{4}$$

оскільки коефіцієнт при аргументі k в функції S_E тут дорівнює одиниці. Вираз отримано з (3) шляхом заміни змінних і являє собою спектральне подання математичної моделі інформаційного локаційного каналу. Як показує практика, частотне уявлення математичної моделі каналу розсіювання систем РАЗ є більш зручним і конструктивним з точки зору вирішення завдань, що розглядаються в роботі, і забезпечує в ряді випадків більш наочні з фізичної точки зору результати.

Графічне подання модулів виразів (2), (3) (4) у трьохмірних координатах називають тілом розсіювання [12]. На рис. 2 представлено тіло розсіювання для зондувальних акустичного та електромагнітного сигналів у вигляді імпульсів з синусоїдальним заповненням та прямокутними формами обвідних.

Як бачимо з рис. 2, амплітуда розсіяного радіосигналу суттєво зменшується при збільшування значення параметру q, саме тому необхідно виконувати підстроювання частоту зондуючого радіосигналу з метою виконання умови Брегга по трасі зондування.

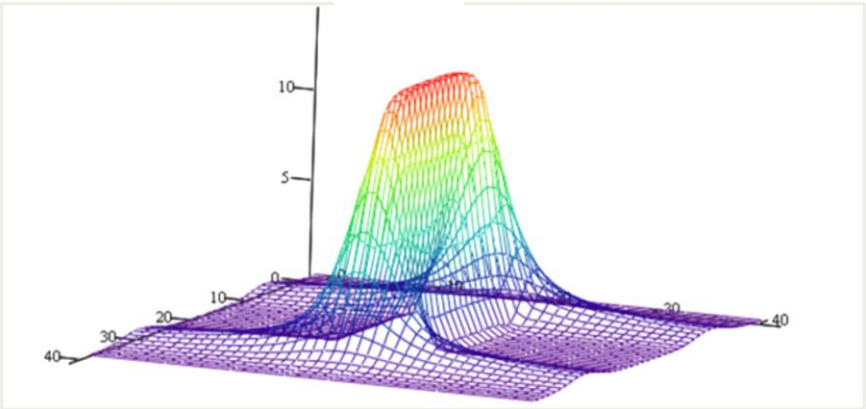


Рис. 2. Тіло розсіювання зондувальних акустичного та електромагнітного імпульсних сигналів з гаусівською та прямокутною формами обвідних

Синтез методу частотної адаптації систем РАЗ

Розглянемо задачу синтезу алгоритму адаптації систем РАЗ шляхом зміни частоти зондувального радіосигналу з метою виконання умови Брегга у міру переміщення випромінюваного акустичного імпульсного сигналу трасою зондування з позицій теорії оптимального управління.

Так як у задачі синтезу алгоритму управління частотою радіосигналу процес, що обурює та призводить до порушень умови Брегга при поширенні акустичного пакета по трасі зондування, а також процес, що викликає похибки вимірювання, розглядаються як випадкові, то дану задачу слід розглядати як завдання стохастичного оптимального управління [30 – 34].

Ступінь розробки питань управління параметрами та структурою радіолокаційних систем нині недостатня. Більшість відомих завдань управління в радіолокаційних комплексах вирішується на основі підходу, який не передбачає достатньої міри формалізації та оптимізації цих завдань, а також чіткої кількісної оцінки якості управління.

У [26 – 29] зазначено, що до найважливіших напрямів подальшого розвитку статистичної теорії вимірювальних радіосистем відноситься побудова прикладної теорії управління станом систем, режимами функціонування, прийому та обробки сигналів, розвиток цієї теорії в рамках завдань вимірювальних комплексів.

Підвищення вимог до функцій управління радіоелектронними системами змушує використовувати при їх проєктуванні адекватний математичний апарат [33]. Таким апаратом є теорія оптимальної динамічної оптимізації, що широко використовується в теорії оптимального управління.

Завдання проєктування оптимальної системи управління у загальному випадку можна сформулювати в такий спосіб [34]: заданий об'єкт чи процес управління; використовуючи деяку інформацію про його стан, потрібно знайти закон управління або керуючу послідовність впливів, що призводять отримання максимуму або мінімуму заданої сукупності критеріїв якості системи.

Завдання управління частотою радіосигналу в РАЗ формулюється наступним чином: на основі послідовного спостереження швидкості акустичного імпульсу при його проходженні по трасі зондування необхідно здійснювати оптимальне, у сенсі обраного критерію якості, підстроювання частоти радіосигналу для виконання умови Брегга. Відповідність між довжинами акустичної та радіохвиль, до якої слід прагнути у процесі керування частотою, зручно представляти графічно у вигляді лінії оптимального управління (рис. 3).

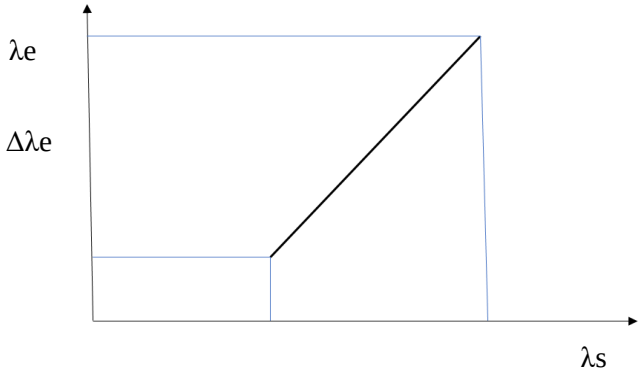


Рис. 3. Лінія оптимального управління

На рис. 3 $\Delta\lambda_s$ – діапазон можливих значень довжин хвиль акустичних коливань; $\Delta\lambda_e$ – подібний діапазон змін довжини хвилі для електромагнітних коливань. Діапазон значень $\Delta\lambda_s$ визначається максимально та мінімально можливими значеннями температури та швидкості поздовжнього вітру.

Системи управління поділяють на системи, що керують об'єктами та системи, що керують спостереженнями [24, 25]. У першому випадку якість функціонування об'єкта визначається управляючими впливами (управліннями), що надходять на його вхід. Інформація про функціонування об'єкта, необхідна для формування управлінь, надходить на вхід системи деякими каналами.

У другому випадку спостерігаються сигнали, які залежать не тільки від інформативних параметрів, у зв'язку з якими повинні бути прийняті певні рішення: виявлення, оцінка, розпізнавання і т.д., а й від параметрів управлінь, що додатково вибираються. Останні вибираються так, щоб забезпечити найкращі якості прийняття рішень.

Очевидно, що завдання управління частотою зондувального радіосигналу для виконання умови Брегга по трасі зондування, іншими словами, для отримання відбитого від пакета сигналу максимальної потужності, відноситься до завдань управління спостереженнями. Характеристики відбитого сигналу залежать не тільки від стану об'єкта зондування – акустичного пакета, а й від параметра управління, що визначає частоту зондуючого сигналу. Частота зондуючого радіосигналу вибирається так, щоб отримати максимально можливу амплітуду відбитого сигналу та забезпечити найкращу якість прийняття рішення – оцінки інформаційних параметрів розсіяного сигналу.

Загальний підхід до вирішення задачі статистичного синтезу алгоритму оптимального управління при використанні моделі в просторі станів полягає в тому, що складається функціонал від функцій стану та управління, визначаються фізичні обмеження на ці функції та знаходяться управління, що мінімізують або максимізують заданий функціонал [31]. Управління повинно бути обрано так, щоб траєкторія зображувальної точки у фазовому просторі задовольняла певному критерію оптимальності. Критерій оптимальності визначається, виходячи з фізичного сенсу розв'язуваної задачі, він повинен бути математично продуктивним і не ускладнювати рішення задачі [34]. В якості такого критерію можуть бути обрані різні показники, наприклад точність, продуктивність та інші характеристики.

Найчастіше критерій якості управління станом системи на деякому кінцевому інтервалі $[0, M]$ задається функціоналом

$$J_M = < \{\sum_{i=1}^{M}([X_i - X_{0i}]^T \mathrm{A}_i[X_i - X_{0i}] + U_i^T B_i U_i)\} >, \quad (5)$$

де X_{0i} – вектор бажаного стану системи в момент часу i; X_i – вектор стану системи; U_i – вектор управління; A_i, B_i – матриці, що визначають вартість похибок стану системи та керуючих зусиль відповідно; знак $<\cdot>$ – означає операцію математичного очікування.

У критерії (5) враховується поведінка як вектору стану, так і вектору управління. Важливість цих двох членів визначається вибором матриць A_i, B_i. Оскільки критерій імовірнісний, використовуються операції математичного очікування, тобто оцінюється якість ансамблю систем з урахуванням усереднення. Зазвичай критерій якості типу (5) інтерпретують як критерій типу «помилка системи плюс керуюче зусилля», тобто він є компромісом між похибкою системи та керуючим зусиллям (енергетичними витратами на управління тощо).

Структура та параметри оптимальної системи управління, отриманої в результаті синтезу, значною мірою визначається критерієм оптимальності. Зауважимо, що формулювання та розв'язання стохастичних завдань мають проводитися на основі ймовірнісних критеріїв, які, на відміну від детермінованих, обов'язково містять операцію статистичного усереднення і тому є складнішими. Часто виявляється, що система оптимальна за деяким критерієм, є квазіоптимальною і за низкою інших критеріїв. Оскільки кожному критерію оптимальності відповідає своя теорія синтезу, вибір конкретного критерію слід проводити з урахуванням розвиненості теорії синтезу та її складності [30].

Аналіз даного конкретного завдання показує, що на керуюче зусилля штраф можна не призначати, оскільки відомі нині технічні засоби можуть забезпечити необхідну швидкість і діапазон перебудови частоти радіопередавача і для цього не потрібно значних енергетичних витрат і витрат іншого роду. При виборі критерію оптимальності слід виходити з вимоги забезпечити мінімум похибки управління, або з вимоги забезпечити виконання більш складніших імовірнісних показників.

У зв'язку з цим в якості критерію оптимальності функціонування пристрою управління частотою зондувального радіосигналу для виконання умови Брегга виберемо точностний критерій. Залежно від вибору відповідних фізичних параметрів можливі два види запису критерію оптимальності

$$J_M =< \{\sum_{i=1}^{M}([\lambda_{ei} - 2\lambda_{si}]^2)\} >, \quad J_M =< \{\sum_{i=1}^{M}([c_{si} - c_{s0i}]^2)\} >. \tag{6}$$

Критерій виду (6) забезпечує мінімум інтегральної дисперсії похибки налаштування, а також мінімум дисперсії налаштування для кожного моменту часу (кожної точки профілю).

Вибраний для вирішення даної задачі квадратичний критерій якості має важливу особливість, яка дозволяє значно спростити розв'язання задачі оптимального синтезу алгоритму оптимального керування частотою радіосигналу. Ця особливість пов'язана з існуванням так званого принципу стохастичної еквівалентності (принцип або теорема розподілу). Даний результат займає дуже важливе місце у завданнях синтезу оптимальних управлінь у лінійних та нелінійних системах при випадкових збуреннях і широко використовується в теорії та на практиці.

Для лінійних систем теорема розподілу формулюється таким чином [34]. Оптимальний регулятор при випадкових гаусових процесах і квадратичному критерії якості являє собою послідовне з'єднання оптимального лінійного фільтра для оцінки вектору стану системи і детермінованого оптимального регулятора. Цей важливий результат дозволяє звести завдання управління до двох послідовно вирішуваних окремих завдань стохастичної фільтрації та детермінованого управління. У цьому матриця передачі зворотного зв'язку системи управління не залежить від стохастичних параметрів завдання, а оптимальний фільтр не залежить від виду критерію якості управління.

В даний час справедливість теореми розподілу доведена також для нелінійних систем та деяких інших критеріїв якості [31]. У нелінійних системах для формування оцінки стану необхідно використовувати нелінійні теорії фільтрації, а власне регулятор, як і раніше, буде детермінованим.

Важливим наслідком теореми розподілу є можливість поєднання результатів досить добре розвиненої теорії фільтрації випадкових процесів та детермінованої теорії оптимального керування. Принцип поділу знаходить широке застосування на практиці при побудові систем управління у різних галузях, зокрема під час управління рухом літальних апаратів і космічних об'єктів [22].

Відповідно до цього пристрій оптимального керування частотою зондувального радіосигналу буде являти собою послідовне з'єднання дискримінатора, оптимального лінійного фільтра та детермінованого регулятора.

Схема оптимального управління частотою зондувального радіосигналу, побудована відповідно до описаного алгоритму, представлена на рис. 4.

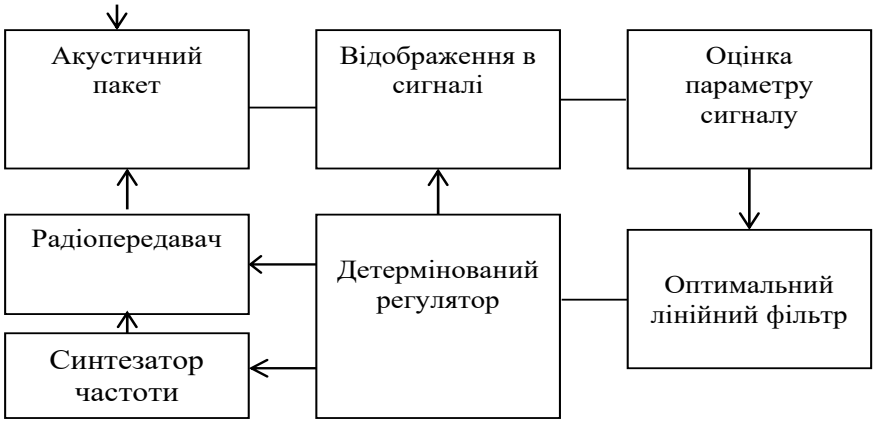


Рис. 4. Схема оптимального управління частотою зондуючого радіосигналу

Основна ідея роботи синтезованого пристрою полягає у наступному. Отримані за результатами вимірювань інформаційного параметра сигналу оцінки вектору стану акустичного пакета (швидкості звуку та його похідних) зазнають в реальному масштабі часу послідовної фільтрації. Дисперсія фільтрованих оцінок суттєво зменшується порівняно з дисперсією вхідних результатів. За уточненими значеннями швидкості звуку в попередніх точках здійснюється оптимальна екстраполяція (прогнозування) його значення в наступну точку профілю. Частота зондуючого радіосигналу в цій точці визначається за екстрапольованим значенням швидкості звуку. Екстраполяція швидкості звуку в наступну точку профілю за відомою передісторією процесу дозволяє значно зменшити динамічну похибку налаштування за умови Брегга.

Обробка швидкості звуку викликана тим, що значення інформаційного параметра в точках профілю знаходяться при різних значеннях частоти зондуючого радіосигналу. А щоб згладжувати відліки інформаційного параметра, їх потрібно «відновити», тобто перерахувати значення даного параметра до однієї частоти зондуючого радіосигналу. Більш доцільною є обробка оцінок швидкості звуку, фільтровані значення яких потім використовуються як для обчислення частоти зондуючого сигналу, так і для визначення метеопараметрів.

При детермінованому управлінні випадкові перешкоди та обурення не враховуються, тому синтез оптимального детермінованого регулятора проведено для критерію якості виду

$$J_M = \{\sum_{i=1}^{M}([\lambda_{ei} - 2\lambda_{si}]^2)\}, \qquad J_M = \{\sum_{i=1}^{M}([c_{si} - c_{s0i}]^2)\}. \tag{7}$$

Функціонали (7) еквівалентні відповідним функціоналам (6), тільки вони не містять операцію статистичного усереднення.

Висновки

1. Процес розсіювання електромагнітних хвиль на акустичному хвильовому пакеті є вузькосмуговим, найбільший рівень розсіяного радіосигналу досягається при виконанні

умови Брегга між несучими частотами акустичного і радіосигналів (або довжинами їх хвиль в атмсфері).

2. Оскільки довжина хвилі акустичних коливань в атмосфері змінюється в залежності від наявних умов, а довжина радіохвилі залишається практично незмінною, то необхідно здійснювати адаптивну зміну частот зондуючих сигналів. Найбільш доцільним є варіант адаптивної зміни частоти зондувального радіосигналу в міру просування звукового пакета атмосфері.

3. Відомі алгоритми частотної адаптації систем РАЗ отримано евристичним шляхом, вони є детермінованими і не задовольняють вимогам, що пред'являються практикою. У статті розглядається синтез методу частотної адаптації систем РАЗ до метеорологічної обстановки, що змінюється, на основі адекватної моделі радіоакустичного інформаційного каналу і з використанням основних досягнень теорії стохастичного оптимального управління.

4. Показано, що у відповідності з теоремою розділення, відомої з теорії оптимального стохастичного управління, метод управління частотою зондувального радіосигналу повинен включати послідовно виконувані операції формування оцінок інформаційного параметра розсіяного радіосигналу, оптимальної лінійної фільтрації отриманих оцінок і детермінованого управління частотою зондуючого радіосигналу.

Список літератури:

1. Bradley S. Atmosphere Acoustic Remote Sensing. Principes andApplication. CRC Press. 2007. 267 p.
2. Kartashov V.M., Tikhonov V.A., Oleinikov V.N. Signal processing in radio electronic systems for remote monitoring of the atmosphere. Kharkiv : KNURE, 2014. 312 p.
3. Карташов В.М. Моделі і методи обробки сигналів систем радіоакустичного і акустичного зондування атмосфери. Харків : ХНУРЕ, 2011. 234 с.
4. Lataitis R.J. Theory and Application of a radio-acoustic sounding system (RASS): NOAA Technical Memorandum ERL WPL-230. Nat. Oceanic and Atmos. Admin. Environ, Res. Labs. Boulder, CO, 1993, 207 p.
5. Smith P. L. Remote measurements of wind velocity by the electromagnetic-acoustic probe. 1. System analysis.1961 // Conf. proc. 5th Annu. convention on military electronics, Wash (D.C.), rep № 419, pp. 43–53.
6. Fetter R. V. Remote measurements of wind velocity by the electromagnetic-acoustic probe. II. Experimental system. 1961 // Conf. proc. 5th Annu. convention on military electronics, Wash (D.C.), rep № 419, pp. 54–59.
7. Atlas D. Indirect probing techniques // Bull.Ainer.Meteorol.Soc. Vol. 43, № 9, pp. 457–466.
8. Marshall I.M., Peterson A.M., and Barnes A.A. Combined Radar-Acoustic Sounding System // Appl. Opt., 1972, v.11, №1. pp. 108–112. DOI: 10.1364/AO.11.000108
9. Ситнік О.В., Карташов В.М. Радіотехнічні системи : навч. посіб. Харків : Сміт, 2009. 448 с.
10. Chandrasekhar Sarma T. V., Narayana Rao D., Furumoto J., and Tsuda T. Development of radio acoustic sounding system (RASS) with Gadanki MST radar – first results // Ann. Geophys., 26, 2008, pp. 2531–2542. https://doi.org/10.5194/angeo-26-2531-2008
11. Alexander S. P., Murphy D. J., Klekociuk A. R., High resolution VHF radar measurements of tropopause structure and variability at Davis, Antarctica (69° S, 78° E). Atmos. Chem. Phys., 13, 2013. pp. 3121–3132. doi:10.5194/acp-13-3121-2013
12. Kartashov V.M. Signal Scattering Functions of Atmospheric Sounding System // Telecommunications and Radio Engineering. 2003, Vol. 59, №7-8-9. P. 88–94.
13. Kartashov V.M. Estimation of Signal Parameters Scattered by an Acoustic Wave Packet // Telecommunications and Radio Engineering. 2004. Vol. 61, №2. P. 125–129.
14. Muradyan P., Richard Coulter R. Radar Wind Profiler (RWP) and Radio Acoustic Sounding System (RASS) Instrument Handbook // March, 2020. Environmental Science Division, Argonne National Laboratory. 20 p. URL: https://www.arm.gov/publications/tech_reports/handbooks/rwp_handbook.pdf.
15. Бабкін С.І., Куценко В.І., Максимова Н.Г. Оцінка похибки двох методик температурного радіоакустичного зондування атмосфери. Експериментальні результати // Радіотехніка. 1988. № 84. С.98–106.
16. Kartashov V., Babkin S., Kartashov A., Pershyn Y. Development of the Atmosphere Radio-Acoustic Sounding Method in Ukraine and in the World in the Period of 1961-2000 // 2023 IEEE 6th International Conference on Information and Telecommunication Technologies and Radio Electronics, UkrMiCo 2023, 13–15 November 2023, Kyiv, Ukraine. pp. 372–376. DOI: 10.1109/UkrMiCo61577.2023.10380339
17. Kartashov V., Oleynikov V., Koryttsev I., Sheiko S., Zubkov O., Babkin S. Processing of Wide Band Acoustic Signals During Detection of Unmanned Aerial Vehicles // 2020 IEEE Ukrainian Microwave Week (UkrMW). Kharkiv, Ukraine, September 21 – 25, 2020. Vol. 1 on 2020 IEEE 12th International Conference on Antenna Theory and Techniques (ICATT). P. 35–39.

18. Oleksandr Sotnikov, Vladimir Kartashov, Oleksandr Tymochko, Oleg Sergiyenko, Vera Tyrsa, Paolo Mercorelli, Wendy Flores-Fuentes. Methods for Ensuring the Accuracy of Radiometric and Optoelectronic Navigation Systems of Flying Robots in a Developed Infrastructure. Chapter 16 // Machine Vision and Navigation. Springer, Cham. P.537–578.
19. Developing and Applying Optoelectronics in Machine Vision / O. Sergiyenko, J.C. Rodriguez-Quiñonez. IGI Global, 2016. 341p.
20. Kartashov V.M., Tikhonov V.A., Voronin V.V. and Tymoshenko L.P. Complex model of random signal in problems of acoustic sounding of atmosphere // Telecommunications and Radio Engineering. 2016. V. 75, Iss. 20. P.1885–1892.
21. Піза Д.М. Теорія і проєктування радіолокаційних систем : навч. посіб.. Запоріжжя : ЗНТУ, 2019. 82 с.
22. Сумик М. М. Основи теорії радіотехнічних систем : навч. посіб. Львів : Львів. політехніка, 2004. 240 с.
23. Радіоелектронні системи : навч. посіб. / Ю.М. Седишев та ін. Харків : ХУПС, 2010. 360 с.
24. Радіоелектронні системи : навч. посіб. / П. Ю. Баранов, В. П. Лавриненко, О. М. Мелешкевич, В. С. Дмитренко. Одеса, 2012. 232 с.
25. Петров В.А., Пилипенко Ю.Л. Радіотехнічні системи. Курсове проєктування : навч. посіб. Харків : ХНУРЕ, 2003. 48 с.
26. Карташов В.М., Тихонов В.А., Воронін В.В., Тимошенко Л.П. Комплексні моделі випадкових сигналів в задачах акустичного зондування атмосфери // Радіотехніка. 2016. Вип. 185. С. 81–86.
27. Vasilchenko A., Kartashov V.M. Analysis of influence exerted by longitudinal Doppler effect upon output signal of sodar antenna array // Telecommunications and Radio Engineering. Vol.66, Iss. 9. P. 841–847. DOI: 10.1615/TelecomRadEng.v66.i9.50.
28. Semenets V. V., Kartashov V.M., Leonidov V. I. Registration of refraction Phenomenon in the Problem of acoustic Sounding of Atmosphere in Airport Zone // Telecommunications and Radio Engineering. 2018. Vol. 77, Iss. 5. P.461–468. DOI: 10.1615/TelecomRadEng.v77.i5.90.
29. Карташов В.М., Тихонов В.А., Воронін В.В. Особливості побудови та застосування комплексних систем дистанційного зондування атмосфери // Радіотехніка. 2016. Вип. 186. С. 184–185.
30. Тютюнник А. Г. Оптимальні і адаптивні системи автоматичного керування : навч. посіб. Житомир : ЖІТІ, 1998. 512 с.
31. Попович М. Г., Ковальчук О. В. Теорія автоматичного керування. Київ : Либідь, 2007. 656 с.
32. Самотокін Б. Б. Лекції з теорії автоматичного керування : навч. посіб. Житомир : ЖІТІ, 2001. 508 с.
33. Бублік Б. Н., Кириченко Н. Ф. Основи теорії управління. К. : Вища шк., 1975. 328 с
34. Іванов А. О. Теорія автоматичного керування. Дніпропетровськ : Нац. гірнич. ун-т, 2003. 250 с.

Надійшла до редколегії 02.06.2024

Відомості про авторів:
Карташов Олександр Володимирович – Харківський національний університет радіоелектроніки, аспірант кафедри медіаінженерії та інформаційних радіоелектронних систем; Україна; e-mail: oleksandr.kartashov@nure.ua, ORCID: https://orcid.org/0000-0002-4618-4787
Кондрашов Ігор Євгенович – Харківський національний університет радіоелектроніки, аспірант кафедри медіаінженерії та інформаційних радіоелектронних систем, Україна, email: igor.kondrashov@nure.ua, ORCID: https://orcid.org/0000-0002-4618-1415

ABSTRACTS РЕФЕРАТИ

SYSTEMS AND METHODS OF INFORMATION PROTECTION
СИСТЕМИ І МЕТОДИ ЗАХИСТУ ІНФОРМАЦІЇ

UDC 004.056

The process of declaring information security profiles / *O.V. Potii, D.Yu. Golubnychiy, Yu.K. Vasiliev, M.V. Yesina* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №217. P. 7 – 22.

The article discusses the process of declaring information security profiles, which is an important aspect of ensuring information security in modern organizations. The main purpose of the declaration is to establish clear requirements and control measures to ensure an appropriate level of protection of information assets against potential threats and vulnerabilities.

The authors of the article analyze the basic and target information security profiles, emphasizing their features, advantages and disadvantages. In particular, the basic security profile is considered as a minimum set of requirements that can be quickly implemented to ensure an initial level of protection. At the same time, the target security profile is aimed at more detailed adaptation of security measures to the specific needs and risks of the organization, which provides a higher level of protection.

The process of declaring security profiles includes several stages, such as assessing the current state of security, identifying threats and vulnerabilities, defining security requirements, developing and implementing security profiles, and regularly monitoring and updating security measures.

The article also discusses the current standards and regulations that govern the process of declaring security profiles, in particular, NIST SP 800-53 rev. 5, ISO/IEC 27001, and ND TZI 3.6-006-21. The article analyzes how these standards can be used to create effective security profiles that meet the specific requirements of organizations of various sizes and industries.

The authors conclude that the process of declaring information security profiles is critical to ensuring an adequate level of protection of information assets. Implementation of clearly defined security profiles allows organizations to take a systematic approach to risk management, reduce the likelihood of security incidents and increase the overall level of information security.

Key words: basic security profile; declaration; classes of security measures; complex information protection systems; estimation; security profile; information security system; target security profile.

2 tabl. 7 fig. Ref: 17 items.

УДК 004.056

Процес декларування профілів безпеки інформації / *О.В. Потій, Д.Ю. Голубничий, Ю.К. Васильєв, М.В. Єсіна* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 217. С. 7 – 22.

Розглядається процес декларування профілів безпеки інформації, що є важливо у забезпеченні інформаційної безпеки в сучасних організаціях. Основною метою декларування є встановлення чітких вимог та контрольних заходів для забезпечення відповідного рівня захисту інформаційних активів від потенційних загроз і вразливостей.

Аналізуються базовий та цільовий профілі безпеки інформації, підкреслюються їхні особливості, переваги та недоліки. Зокрема, базовий профіль безпеки розглядається як мінімальний набір вимог, який може бути швидко реалізований для забезпечення початкового рівня захисту. Водночас, цільовий профіль безпеки спрямований на більш детальну адаптацію заходів безпеки до специфічних потреб та ризиків організації, що забезпечує більш високий рівень захисту.

Процес декларування профілів безпеки включає кілька етапів, таких як оцінка поточного стану безпеки, ідентифікація загроз та вразливостей, визначення вимог до безпеки, розробка та впровадження профілів безпеки, а також регулярний моніторинг і оновлення заходів безпеки.

Обговорюються сучасні стандарти та нормативні документи, які регулюють процес декларування профілів безпеки, зокрема, NIST SP 800-53 rev. 5, ISO/IEC 27001 та НД ТЗІ 3.6-006-21. Аналізується, як ці стандарти можуть бути використані для формування ефективних профілів безпеки, що відповідають специфічним вимогам організацій різних розмірів та галузей.

Автори роблять висновок, що процес декларування профілів безпеки інформації є критично важливим для забезпечення належного рівня захисту інформаційних активів. Впровадження чітко визначених профілів безпеки дозволяє організаціям систематично підходити до управління ризиками, знижувати ймовірність інцидентів безпеки та підвищувати загальний рівень інформаційної безпеки.

Ключові слова: базовий профіль безпеки; декларація; класи заходів захисту; комплексні системи захисту інформації; оцінювання; профіль безпеки; система безпеки інформації; цільовий профіль безпеки.

Табл. 2. Іл. 7. Бібліогр.: 17 назв.

UDC 004.056.5

Justification of methods for calculating and analyzing the properties of pseudorandom and random sequences based on DNA / *Ya.A. Derevianko, M.V. Yesina, D.Yu. Gorbenko* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №217. P. 23 – 38.

An integral requirement for modern information systems is to provide users with services such as confidentiality, integrity, availability, and irrefutability. The quality of such services directly depends on cryptographic transformations, an important component of most of which is randomness. Therefore, the generation of pseudorandom and random sequences is one of the most relevant and important tasks in cryptography. Such sequences are generated based on physical and non-physical noise sources. Our previous studies indicate the theoretical possibility to use DNA as a noise source and, accordingly, as a source of random sequences.

Any noise source "contains randomness", i.e. it has a certain amount of entropy, but a sample from such a source will not always have good properties. That is why there is a need for tools that can obtain sequences with good randomness properties of samples from the DS sequences with good randomness properties of samples from the noise source (by, for example, some kind of enhancement). Such sequences should satisfy the necessary conditions: be statistically indistinguishable, uniform, etc. NIST-approved DRBG designs can be used to guarantee this. Such constructions are most often based on strong crypto-primitives, such as hashes, HMACs, or block or stream ciphers in the required modes.

This work is devoted to newly developed methods for obtaining pseudorandom and random sequences based on DNA sequences using the national standard for block symmetric encryption DSTU 7624:2014 in the counter (CTR) operation mode, as well as methods for comparing sequences (both DNA and binary).

The work essentially opens the topic of obtaining random sequences based on DNA, since previous studies of DNA in cryptography have focused on the use of DNA in encryption and steganography. The paper presents the results of solving such issues as the development of methods for obtaining DNA-based sequences, the evaluation of statistical and stochastic properties of such sequences, and the evaluation of similarity based on k-mer and MinHash distances.

The results obtained indicate the prospects and relevance of further research in this area.

Key words: DNA; random sequences; extractors; symmetric encryption; post-quantum standards; statistical testing; hashing; similarity assessment.

7 tabl. 9 fig. Ref: 15 items.

УДК 004.056.5

Обґрунтування методів обчислення та аналіз властивостей псевдовипадкових та випадкових послідовностей на основі ДНК / *Я.А. Дерев'янко, М.В. Єсіна, Д.Ю. Горбенко* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 217. С. 23 – 38.

Для сучасних інформаційних систем невід'ємною вимогою є надання користувачам таких послуг як конфіденційність, цілісність, доступність та неспростовність. Якість таких послуг напряму залежить від криптографічних перетворень, важливою складовою більшості з яких є випадковість. Тому генерування псевдовипадкових (ПВП) та випадкових (ВП) послідовностей є однією з актуальних та важливих задач криптографії. Такі послідовності генеруються на основі фізичних та нефізичних джерел шуму. Наші попередні дослідження вказують на теоретичну можливість використання у якості джерела шуму (ДШ) та відповідно джерела ВП ДНК.

Будь-яке ДШ «містить випадковість», тобто має певну ентропію, але вибірка з такого джерела не завжди матиме хороші властивості. Саме тому існує необхідність у інструментах, які можуть отримати послідовності з хорошими властивостями випадковості з вибірок з ДШ (шляхом, наприклад, певного покращення). Такі послідовності повинні задовольняти необхідні умови: бути статистично нерозрізнюваними, рівномірними, тощо. Щоб гарантувати це, можна використовувати, затверджені NIST конструкції DRBG. Такі конструкції найчастіше базуються на стійких криптопримітивах, наприклад гешах, HMAC або блокових чи потокових шифрах в необхідних режимах.

Робота присвячена новим розробленим методам отримання ПВП та ВП на основі послідовностей ДНК з використанням національного стандарту блокового симетричного шифрування ДСТУ 7624:2014 у режимі гамування, а також методам порівняння послідовностей (як ДНК, так і двійкових).

Робота відкриває тему отримання саме випадкових послідовностей на основі ДНК, оскільки попередні дослідження ДНК у криптографії були зосереджені на використанні ДНК у шифруванні та стеганографії. Надано результати вирішення таких питань як розробка методів отримання послідовностей на основі ДНК, оцінка статистичних властивостей та стохастичних показників таких послідовностей, а також оцінка подібності на основі k-мер та MinHash відстаней.

Отримані результати вказують на перспективність та актуальність подальших досліджень у даному напрямку.

Ключові слова: ДНК; випадкові послідовності; екстрактори; симетричне шифрування; постквантові стандарти; статистичне тестування; гешування; оцінка подібності

Табл. 7. Іл. 9. Бібліогр.: 15 назв.

UDC 004.05

Review of existing models and basic zero trust principles / *V.I. Yesin, V.V. Vilihura, D.Y. Uzlov* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №217. P. 39 – 54.

Ensuring the information security of an enterprise is quite a complex task. This is due to the multifaceted nature of IT infrastructure and applications, the breadth and intensity of user access, the excessive openness of most corporate networks, and several other factors. In these conditions, the concept of zero trust is increasingly being considered as the most preferable solution to the problem of ensuring the security of enterprises, organizations, institutions. The basic idea of the concept of zero trust is that there are no areas that are trustworthy. However, despite the popularization of the zero trust concept and the obvious security benefits of its application in enterprises, there are certain difficulties in its implementation. In particular, planning to bring the infrastructure into compliance with the zero-trust principles cannot be accomplished partially or as part of minor modifications to the relevant information systems. It is necessary to reorganize the information infrastructure as a whole, as well as to integrate all aspects that ensure the security of enterprise activities, so that the zero-trust principles show their effectiveness. On the other hand, today there is a problem associated with a certain lack of awareness about the zero-trust approach (about its theoretical and practical potential) for choosing the right solution. This paper is precisely aimed at solving this problem by summarizing existing research and the experience of various international companies that are implementing this approach in practice. It briefly discusses models and key zero-trust principles proposed by renowned international organizations and companies that will help make sense of a fundamental shift in the approach to information security, cybersecurity.

Key words: zero trust; models and principles of zero trust; cybersecurity; information security.

3 tabl. 3 fig. Ref: 41 items.

УДК 004.05

Огляд існуючих моделей та основних принципів нульової довіри / *В.І. Єсін, В.В. Вілігура, Д.Ю. Узлов* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 217. С. 39 – 54.

Забезпечити інформаційну безпеку підприємства є досить складним завданням. Це пояснюється багатогранністю ІТ-інфраструктури та застосунків, широтою та інтенсивністю доступу користувачів, надмірною відкритістю більшості корпоративних мереж та деяких інших факторів. В цих умовах концепція нульової довіри дедалі частіше розглядається як найбільш переважне вирішення проблеми забезпечення безпеки підприємств, організацій, установ. Основна ідея концепції нульової довіри полягає в тому, що не існує областей, які заслуговують на довіру. Однак, незважаючи на популяризацію концепції нульової довіри та очевидні переваги у сфері безпеки від її застосування, на підприємствах виникають певні складнощі щодо її реалізації. Зокрема, планування приведення інфраструктури у відповідність до принципів нульової довіри неможливо здійснити частково або в рамках незначного доопрацювання відповідних інформаційних систем. Необхідна реорганізація інформаційної інфраструктури в цілому, а також інтеграція всіх аспектів, що забезпечують безпеку діяльності підприємства, щоб принципи нульової довіри показали свою ефективність. З іншого боку, існує проблема, пов'язана з певним дефіцитом поінформованості про підхід нульової довіри (про його теоретичний та практичний потенціал) для вибору правильного рішення. Стаття націлена на вирішення цієї проблеми шляхом узагальнення наявних досліджень та досвіду різних міжнародних компаній, які впроваджують даний підхід на практиці. Стисло розглядаються моделі та ключові принципи нульової довіри, запропоновані відомими міжнародними організаціями та компаніями, які допоможуть розібратися у фундаментальному зрушенні у підході до інформаційної безпеки, кібербезпеки.

Ключові слова: нульова довіра; моделі і принципи нульової довіри; кібербезпека; інформаційна безпека.

Табл. 3. Іл. 3. Бібліогр.: 41 назв.

UDC 004.056.5

Using machine learning to classify DOS/DDOS attacks / *M.S. Kavetskyi, O.V. Sievierinov, R.Y. Gvozdov, A.O. Smirnov* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №217. P. 55 – 63.

The relevance of this work is manifested in the need to detect and counteract DOS/DDOS attacks, which pose a serious threat to modern information systems. These cyberattacks lead to significant economic losses and disruptions in the operation of network services. The aim of the work is to confirm the hypothesis that the decision tree method performs better for detecting DOS/DDOS attacks under certain conditions.

A comparison of decision tree methods with other machine learning methods (RF, SVM, KNN, ANN, NB, SGBoost) was conducted based on the CSICIDS2017 dataset. Decision trees have shown significant improvements in attack detection accuracy through optimal hyperparameter tuning and dataset selection.

Key words: machine learning; decision trees; CSICIDS2017; RF; SVM; KNN; ANN; NB; SGBoost; DOS; DDOS; classifier.

1 tabl. 1 fig. Ref: 6 items.

УДК 004.056.5

Використання машинного навчання для класифікації атак типу DOS/DDOS / *М.С. Кавецький, О.В. Сєвєрінов, Р.Ю. Гвоздьов, А.О. Смірнов* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 217. С. 55 – 63.

Актуальність роботи полягає у необхідності виявлення та протидії атакам типу DOS/DDOS, що є серйозною загрозою для сучасних інформаційних систем. Ці кібератаки призводять до значних економічних збитків та перерв у роботі мережевих сервісів. Мета роботи – підтвердити гіпотезу, що метод дерев прийняття рішень краще працює для виявлення атак типу DOS/DDOS за певних умов.

Проведено порівняння методів дерев прийняття рішень з іншими методами машинного навчання (RF, SVM, KNN, ANN, NB, SGBoost) на основі датасету CSICIDS2017. Дерева прийняття рішень показали значні покращення у точності виявлення атак завдяки оптимальному налаштуванню гіперпараметрів та відбору датасету.

Ключові слова: машинне навчання; дерева прийняття рішень; CSICIDS2017; RF; SVM; KNN; ANN; NB; SGBoost; DOS; DDOS; класифікатор

Табл. 1. Іл. 1. Бібліогр.: 6 назв.

UDC 004.056.5

Analysis of methods for bypassing modern EDR endpoint protection systems / *K.M. Shulika, D.S. Balagura, Z.M. Sydorenko* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №217. P. 64 – 68.

The purpose of the article is to review and analyze the methods of bypassing complex solutions for endpoint protection (EDR). The article highlights and describes the salient features of each of the EDR bypass methods and provides recommendations for countering them. EDR (Endpoint Detection and Response) is a type of cross-platform software currently most commonly used for event monitoring, security incident generation and formalization, and incident response. For each method and tool, example of it's useand advantages and disadvantages are described. The article will be useful for cybersecurity analysts who want to deepen their knowledge of EDR and strengthen endpoint security. It will provide readers with an insight into EDR bypass techniques and help them apply recommendations to reduce the risks of EDR bypass during cyber ttacks.

Key words: endpoint protection system; EDR; SOC; AMSI bypass; unhooking; reflective DLL

1 tabl. 1 fig. Ref: 11 items.

УДК 004.056.5

Аналіз методів обходу сучасних систем захисту кінцевих точок EDR / *К.М. Шуліка, Д.С. Балагура, З.М. Сидоренко* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 217. С. 64 – 68.

Метою статті є огляд та аналіз методів обходу комплексних рішень для захисту кінцевих точок (EDR). Виділяються та описуються визначні риси кожного з методів обходу EDR та наводяться рекомендації з протидії ним. EDR (Endpoint Detection and Response) є типом кросплатформенного програмного забезпечення, що наразі найчастіше використовується для моніторингу подій, формування та формалізації інцидентів безпеки та реагування на інциденти на кінцевих точках. Для кожного методу та інструменту наводиться приклад його використання та описуються переваги та недоліки. Стаття буде корисна аналітикам в сфері кібербезпеки, які хочуть поглибити знання про EDR та посилити безпеку кінцевих точок. Вона надасть читачам уявлення про методи обходу EDR та допоможе їм застосувати рекомендації для зменшення ризиків обходу EDR під час кібератак.

Ключові слова: система захисту кінцевих точок; EDR; SOC; обхід AMSI; метод зняття з крючка; рефлективна DLL

Табл. 1. Іл. 1. Бібліогр.: 11 назв.

UDC 004.056.5

Evaluation and comparison of lattice-based digital signature of the "Digital Signature Schemes" PQC NIST competition / *Yu.I. Gorbenko, Ye.V. Ostrianska* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №217. P. 69 – 78.

Over the past decade, post-quantum cryptography has reached a tipping point; institutional bodies and stakeholders have initiated standardization and deployment, and various projects have achieved a reasonably high level of progress and even deployment and implementation. In July 2022, at the end of Round 3 of the NIST's PQC competition, 3 candidates were proposed for the NIST standardization for post-quantum digital signatures scheme: one signature scheme based on MLWE (Crystals-Dilithium), one signature based on NTRU (Falcon), and one signature based on hash (Sphincs+). Although the performance profiles and "black-box" security of these schemes are well understood, resistance to side-channel attacks remains a weak point for all of them. After that, the NIST announced that the PQC standardization process is continuing with a fourth round, with the following KEMs still under consideration: BIKE, Classic McEliece, HQC, and SIKE. However, there are no candidates of digital signature schemes left for consideration. As such, the NIST has issued a call for additional digital signature proposals to be considered in the PQC standardization process. Acceptance of documents ended on June 1, 2023. As a result, 40 candidates were selected for the role of DS standard, namely: 6 DS algorithms based on codes, one DS algorithm based on isogenies, 7 DS algorithms based on lattice operations, 7 candidates for the role of DS algorithm based on the MPC method -in-the-Head and 10 algorithms based on multivariate transformations, 4 DS schemes were selected based on symmetric cryptographic transformations, and 5 more candidates based on other types of cryptographic transformations. The NIST is primarily interested in addi-

tional general purpose signature schemes that are not based on structured lattices. For certain applications, such as certificate transparency, the NIST may also be interested in signature schemes that have short signatures and fast verification. The NIST is open to receiving additional materials based on structured lattices, but intends to diversify post-quantum signature standards. Therefore, any structured array-based signature proposal would need to significantly outperform CRYSTALS-Dilithium and FALCON in relevant applications and/or provide significant additional security properties to be considered for standardization. Thus, the purpose of this paper is to analyze, evaluate, and compare digital signature algorithms based on lattice cryptography, an additional PQC NIST competition, and compare them with already standardized lattice-based DS mechanisms, such as CRYSTALS-Dilithium and FALCON.

Key words: post-quantum cryptography; signature scheme; digital signature; lattice-based cryptography; NIST PQC.

5 tabl. Ref: 21 items.

УДК 004.056.5

Оцінка та порівняння криптоперетворень типу ЕП на основі криптографії на решітках конкурсу NIST США «Digital Signature Schemes» / *Ю.І. Горбенко, Є.В. Острянська* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 217. С. 69 – 78.

За останнє десятиліття постквантова криптографія досягла переломного моменту; інституційні органи та зацікавлені сторони ініціювали стандартизацію та розгортання, і різноманітні проєкти досягли достатньо високого рівня прогресу та, навіть розгортання та впровадження. У липні 2022 р., наприкінці 3-го раунду конкурсу PQC NIST, щодо постквантових цифрових підписів було запропоновано три кандидати на стандартизацію NIST: один підпис на основі MLWE (Crystals-Dilithium), один підпис на основі NTRU (Falcon) і один підпис на основі гешу (Sphincs+). Хоча профілі ефективності та безпека «чорної скриньки» цих схем добре зрозумілі, стійкість до атак із бічних каналів залишається слабким місцем для всіх них. Після чого NIST оголосив, що процес стандартизації PQC продовжується четвертим раундом, при цьому наступні KEM все ще знаходяться на розгляді: BIKE, Classic McEliece, HQC і SIKE. Однак на розгляді не залишилося жодного кандидата на цифровий підпис. Таким чином, NIST опублікував заклик до додаткових пропозицій щодо цифрового підпису, які слід розглянути в процесі стандартизації PQC. Прийом документів завершився 1 червня 2023 р. В результаті було обрано 40 кандидатів на роль стандарту ЕП, а саме: 6 алгоритмів ЕП на основі кодів, один алгоритм ЕП на основі ізогеній, 7 алгоритмів ЕП, в основі яких лежать операції на решітках, 7 кандидатів на роль алгоритму ЕП на основі методу MPC-in-the-Head та 10 алгоритмів, в основі яких лежать багатоваріативні перетворення, на основі симетричних криптоперетворень було обрано чотири схеми ЕП, та ще п’ять кандидатів, що базуються на інших видах криптографічних перетворень. NIST насамперед зацікавлений у додаткових схемах підписів загального призначення, які не базуються на структурованих решітках. Для певних застосувань, таких як прозорість сертифікатів, NIST також може бути зацікавлений у схемах підписів, які мають короткі підписи та швидку перевірку. NIST відкритий для отримання додаткових матеріалів на основі структурованих решіток, але має намір урізноманітнити стандарти постквантових підписів. Таким чином, будь-яка пропозиція підпису на основі структурованої решітки повинна буде значно перевершувати CRYSTALS-Dilithium і FALCON у відповідних додатках і/або забезпечувати значні додаткові властивості безпеки, які будуть розглянуті для стандартизації. Мета статті - аналіз, оцінка та порівняння алгоритмів ЕП, в основі яких лежить криптографія на решітках, додаткового конкурсу NIST США та їх порівняння з вже стандартизованими механізмами ЕП на решітках, таких як CRYSTALS-Dilithium і FALCON.

Ключові слова: постквантова криптографія; схема підпису; електронний підпис; криптографія на решітках; NIST PQC.

Табл. 5. Бібліогр.: 21 назв.

UDC 004.056.55

Assessing the influence of the algebraic structure of q-ary lattices on the complexity of cryptanalysis of problems on lattices / *S.O. Kandii, I.D. Gorbenko* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №217. P. 79 – 99.

The work is devoted to the influence of the algebraic structure of q-ary lattices on the complexity of cryptanalysis of cryptographic problems on lattices. The root-mean-square errors for existing lattice reduction models are obtained. It is shown that the deterministic Albrecht-Lee simulator shows the smallest mean square deviations on small lattices. Collapsibility estimates for nesting and decoding attacks on the algebraic structure of q-ary lattices have been found. A new method for selecting attack parameters has been proposed for the decoding attack. It is shown that a decoding attack with such a choice of parameters can be at the expense of nesting attacks. To distribute the power of the secret vector in decoding attacks, which differ from the normal one, the approximation method is used by the normal division, which minimizes the Kolmogorov-Smirnov equation and calculates the optimal values for some distributions. There are no parameters for any divisions. For the SIS problem, a new approach to safety assessment has been proposed, which takes into account the possibility of various Euclidean norms. Based on the new approach, estimates of the SIS problem for Crystals-Dilithium electronic signature schemes have been calculated.

Key words: NTRU; SIS; LWE; cryptanalysis; lattice cryptography; Crystals-Dilithium.

2 tabl. 16 fig. Ref: 23 items.

УДК 004.056.55

Оцінка впливу алгебраїчної структури q-арних решіток на складність криптоаналізу проблем на решітках / *С.О. Кандій, І.Д. Горбенко* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 217. С. 79 – 99.

Робота присвячена впливу алгебраїчної структури q-арних решіток на складність криптоаналізу криптографічних проблем на решітках. Отримано середньоквадратичні похибки для існуючих моделей редукції решіток. Показано, що детермінований симулятор Альбрехта-Лі показує найменші середньоквадратичні похибки на решітках малої розмірності. Знайдено оцінки складності для атак вкладення та декодування з врахуванням алгебраїчної структури q-арних решіток. Для атаки декодування запропоновано новий метод вибору параметрів атаки. Показано, що атака декодування при такому виборі параметрів може бути кращою за атаки вкладення. Для розподілів ймовірностей таємного вектора у атаках декодування, що відрізняються від нормального, запропонований метод апроксимації нормальним розподілом, що мінімізує відстань Колмогорова–Смірнова та обчислено оптимальні значення параметрів для деяких розподілів. Для проблеми SIS запропоновано новий підхід до оцінки безпеки, що враховує можливість різних евклідових норм. У межах нового підходу обчислено оцінки проблеми SIS для схеми електронного підпису Crystals-Dilithium.

Ключові слова: NTRU; SIS; LWE; cryptanalysis; lattice cryptography; Crystals-Dilithium.

Табл. 2. Іл. 16. Бібліогр.: 23 назв.

UDC **621.391:519.2**

Modified genetic algorithms for generating S-boxes with high nonlinearity / *O. Kuznetsov, M. Poluyanenko, D. Prokopovych-Tkachenko, Y. Kotukh, V. Liubchak* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. № 217. P.100 – 109.

This article discusses the use of modified genetic algorithms to generate S-boxes with high nonlinearity, which is critical for ensuring the security of cryptographic algorithms. S-boxes play a key role in providing resistance to cryptanalysis, in particular to linear and differential cryptanalysis attacks. In the course of the study, a series of experiments were conducted using a genetic algorithm modified by additional storage of the current population and the use of selection. This approach has significantly improved the efficiency of the algorithm compared to the classical genetic algorithm. The best results were achieved with the minimum number of instances in the population and the optimal number of mutations. It was found that the algorithm finds S-blocks with a probability of 99% with a nonlinearity of 104, which demonstrates its high efficiency. The results of the study showed that the modified algorithm is able to provide stable generation of S-boxes with the required attack resistance properties. The proposed method also proved to be flexible in parameter settings, which allows it to be adapted for various cryptographic applications. The paper also discusses the possibilities of further improving the algorithm, including the study of other mutation and selection methods, as well as parameter optimization to achieve even better results. Additionally, the possibility of using distributed computing to increase the speed of S-boxes generation is considered. The results of the study indicate the prospects of the proposed approach for creating attack-resistant cryptographic systems.

This study considered the use of genetic algorithms to generate S-boxes with high nonlinearity. The proposed approach was based on the use of an objective function with optimal parameters, which allowed achieving high efficiency in generating bioactive S-blocks.

The proposed algorithm showed high stability and efficiency in generating S-boxes with nonlinearity, which is confirmed by a 99% probability of achieving the goal. These results indicate the prospects of using genetic algorithms in cryptographic applications that require high resistance to attacks.

In future research, it is planned to further improve the algorithm by exploring other methods of mutation and selection, as well as optimizing parameters to achieve even better results. In addition, it is possible to use distributed computing to further speed up the process of generating S-boxes.

Key words: genetic algorithm; S-box; cryptography; nonlinearity; Walsh-Hadamard Spectrum; selection; mutation; cryptanalysis; optimization; data security.

Табл. 2. Іл. 7. Бібліогр.: 22 назв.

УДК 621.391:519.2

Модифіковані генетичні алгоритми для генерації S-boxes з високою нелінійністю / *О.О. Кузнецов, М. О. Полуяненко, Д.І. Прокопович-Ткаченко, Є.В. Котух, В.О. Любчак* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 217. С. 100 –109.

Розглядається застосування модифікованих генетичних алгоритмів для генерації S-boxes з високою нелінійністю, що є критично важливим для забезпечення безпеки криптографічних алгоритмів. S-boxes відіграють ключову роль у забезпеченні стійкості до криптоаналізу, зокрема до атак методом лінійного та диференціального криптоаналізу. Проведено серію експериментів з використанням генетичного алгоритму, модифікованого додатковим зберіганням поточної популяції та застосуванням селекції. Такий підхід дозволив значно покращити ефективність алгоритму порівняно з класичним генетичним алгоритмом. Найкращі результати досягнуто при мінімальній кількості екземплярів у популяції та оптимальному числі мутацій. Встановлено, що алгоритм з ймовірністю 99 % знаходить S-блоки з нелінійністю 104, що демонструє його високу ефективність. Результати дослідження показали, що модифікований алгоритм здатний забезпечити стабільну генерацію S-boxes з необхідними властивостями стійкості до атак. Запропонований метод виявився гнучким у налаштуванні параметрів,

що дозволяє адаптувати його для різних криптографічних застосувань. Обговорюються можливості подальшого вдосконалення алгоритму, включаючи дослідження інших методів мутацій та селекції, а також оптимізацію параметрів для досягнення ще кращих результатів. Додатково розглядається можливість використання розподілених обчислень для підвищення швидкості генерації S-boxes. Результати дослідження свідчать про перспективність запропонованого підходу для створення стійких до атак криптографічних систем.

Ключові слова: генетичний алгоритм; S-box; криптографія; нелінійність; Walsh–Hadamard Spectrum; селекція; мутація; криптоаналіз; оптимізація; безпека даних.

Табл. 2. Іл. 7. Бібліогр.: 22 назв.

MEANS OF TELECOMMUNICATIONS
ЗАСОБИ ТЕЛЕКОМУНІКАЦІЙ

UDC 621.396

Support of resourses redistribution in NB-IoT LTE networks / *O.I. Kadatskaya, S.A. Saburova //* Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №217. P. 110 – 116.

The NB-IoT is expected to be used with the deployed LTE network as well as future 5G networks. The rapid development of the IoT concept has entailed the need to provide wireless communication to a huge number of devices included in the infrastructure. As part of the 5G NR standard for such devices, Massive Machine Communications (m MTC) technology is focused on optimizing the use of network resources to support a large number of stable connections per unit area.

The Narrowband Internet of Things (NB-IoT) has been analyzed to enable the connectivity of a wide range of new IoT devices and services to the mobile network. The NB-IoT is also shown to be designed for fixed devices with low data transmission and low consumption, leading to an increase in the number of interconnected devices. In turn, standard NB-IoT modules attempting to simultaneously request radio channel resources for uplink data transmission may suffer from random access preamble collision. The proposed model describes the macro- and microcells of an NB-IoT LTE cluster.

An increase in the efficiency of using the bandwidth of networks based on macro- and microcells with a high concentration of users of the NB-IoT LTE networks is shown. Numerical results of an analysis of identifying factors affecting system performance are presented. A linear increase in throughput has been revealed depending on the throughput of the macrocell when using the shared resource of the macrocell for a microcell of equal size without prior repacking of channels when servicing moving subscribers. In the absence of moving subscribers, the additional load is serviced, and the efficiency of using a macro cell increases almost 3 times. In addition, repacking channels significantly increases system throughput.

Key words: narrowband; Internet; things; efficiency; macrocells: microcells; resource; model; throughput; repackaging; channel; network.

3 fig. Ref: 10 items.

УДК 621.396

Підтримка перерозподілу ресурсів у мережах NB-IoT LTE / *О.Й. Кадацька, С.О. Сабурова* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 217. С. 110 – 116.

Очікується, що NB-IoT використовуватиметься з розгорнутою нині мережею LTE, а також з майбутніми мережами 5G. Стрімкий розвиток концепції IoT спричинив необхідність забезпечення безпроводового зв'язку величезної кількості пристроїв, що є частиною інфраструктури. У рамках стандарту 5G NR для таких пристроїв масовий зв'язок машинного типу (mMTC) орієнтовано на оптимізацію використання мережевих ресурсів для підтримки великої кількості стабільних з'єднань на одиницю площі.

Проведено аналіз вузькосмугового Інтернету речей (NB-IoT), який дозволяє використовувати широкий спектр нових пристроїв та послуг IoT, підключених до мобільної мережі. Також показано, що NB-IoT призначено для стаціонарних пристроїв з низькою передачею даних і низьким споживанням, що призводить до збільшення кількості пристроїв, які підключаються один до одного. У свою чергу, масові модулі NB-IoT, які намагаються одночасно запросити ресурси радіоканалу для передачі даних по висхідній лінії зв'язку, можуть постраждати від колізії преамбули довільного доступу. Запропонована модель описує макро- та мікроосередки NB-IoT LTE кластера.

Показано підвищення ефективності використання пропускної спроможності мереж на базі макро- та мікросот для високої концентрації користувачів мереж NB-IoT LTE. Наведено числові результати аналізу виявлення чинників, які впливають на продуктивність системи. Виявлено лінійне збільшення пропускної спроможності залежно від ємності макроосередка при використанні загального ресурсу макроосередку для мікроосередки рівного розміру без попереднього переупаковування каналів при обслуговуванні абонентів, що рухаються. За відсутності абонентів, що рухаються, забезпечується обслуговування додаткового навантаження, ефективність використання макроосередку зростає майже в три рази. Крім того, переупаковка каналів істотно збільшує пропускну здатність системи.

Ключові слова: вузькосмуговий; Інтернет; речі; ефективність; макроосередок: мікроосередок; ресурс; модель; пропускна спроможність; переупаковка; канал; мережа.

Іл. 3. Бібліогр.: 10 назв.

UDC 621.396.004

Study of data replication process using Raft replication algorithm to maintain consistency in server cluster / *L.O. Tokar, V.Y. Tsyliuryk, V.V. Solodilov //* Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №217. P. 117 – 127.

The article considers the issues of ensuring availability and fault tolerance of a server cluster. The analysis of methods and technologies for increasing system reliability and improving performance, such as data replication, automatic switching between servers and fast recovery after failures, is carried out. It is indicated that the key element in modern distributed systems is replication-based clustering. The types of replication are analyzed. It is shown that during the operation of the system, there is a choice between the stability of the received data and the speed or scaling limitation.

It is shown that the Raft algorithm is one of the reliable mechanisms with a high level of availability for achieving consensus and data management. An analysis of the literature on the use of the Raft consensus algorithm is carried out.

The replication process using the Raft algorithm is analyzed. It is substantiated that it is possible to ensure strong consistency and high availability of the cluster using the management and control mechanism on the Kubernetes platform while maintaining cluster configuration synchronization.

The Raft consensus process is studied using the M/M/s QS model. The exponential distribution of time between client requests and servicing with a certain number of servers is considered. The program code for the mathematical model in Python has been developed.

Modeling has been performed in the Visual code environment in Python. The following parameters have been studied: the number of servers that operate at a certain average speed, the expected time of a client request in the system, and the probability of message delay. It has been proven that as the arrival speed increases, the expected time of a client request increases. The analysis provides an understanding of how the Raft algorithm works in different contexts and is applicable to optimizing system design processes.

Key words: server; cluster; fault tolerance; availability; replication; consensus; Kubernetes; Raft.

9 fig. Ref: 15 items.

УДК 621.396.004

Дослідження процесу реплікації даних за допомогою алгоритма реплікації Raft для підтримки узгодженості в кластері серверів / *Л.О. Токар, В.Є. Цилюрик, В.В. Солоділов* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 217. С. 117 – 127.

Розглянуто питання забезпечення доступності та відмовостійкості кластера серверів. Проаналізовано методи та технології підвищення надійності систем й покращення продуктивності, таких як реплікація даних, автоматичне перемикання між серверами та швидке відновлення після відмов. Зазначено, що ключовим елементом у сучасних розподілених системах є кластеризація на основі реплікації. Проаналізовано типи реплікації. Показано, що в процесі роботи системи є вибір між стійкістю отриманих даних та обмеженням в швидкості чи в масштабуванні.

Показано, що одним з надійних механізмів з високим рівнем доступності для досягнення консенсусу та управління даними є алгоритм Raft. Проведено аналіз літератури з використання алгоритму консенсусу Raft.

Проаналізовано процес реплікації з використанням алгоритму Raft. Обґрунтовано, що забезпечити сильну консистентність та високу доступність кластера можливо з використанням механізму управління та контролю на платформі Kubernetes із збереженням синхронізації конфігурації кластера.

Проведено дослідження процесу консенсусу Raft з використанням моделі СМО M/M/c. Розглянуто експоненційний розподіл часів між запитами клієнтів та обслуговуванням з певною кількістю серверів. Розроблено код програми для математичної моделі на мові Python.

Проведено моделювання в середовищі Visual code на Python. Досліджено параметри: кількість серверів, що працюють з певною середньою швидкістю, очікуваний час запиту клієнта у системі, ймовірність затримки повідомлення. Доведено, що по мірі збільшення швидкості прибуття очікуваний час запиту клієнта зростає. Аналіз дає розуміння того, як алгоритм Raft працює в різних контекстах і може бути застосований для оптимізації процесів проєктування систем.

Ключові слова: сервер; кластер; відмовостійкість; доступність; реплікація; консенсус; Kubernetes; Raft.

Іл. 9. Бібліогр.: 15 назв.

RADIO ENGINEERING DEVICES
РАДІОТЕХНІЧНІ ПРИСТРОЇ

UDC 615.472

Software and hardware complex based on the STM32F407VG microcontroller for the study of vibrations with the LIS3DSH accelerometer / *V.V. Semenets, A.B. Grigoriev* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №217. P. 128 – 132.

Commonly used color models are reviewed and analyzed, namely the RGB model and subtractive models, including CMY and CMYK. The characteristics of colorimeters based on the use of RGB, CMY and CMYK models have been studied. A comparative analysis of the advantages and disadvantages of RGB and CMYK models was carried out.

The developed digital colorimeter allows for express control of color characteristics with high metrological accuracy and does not require a high level of qualification of service personnel. This device reproduces the perception of color by human vision, and its use is intended for both people with normal and impaired vision. The great advantage of this device is its mobility, the speed of measurements and the absence of the need for a specialized laboratory and highly qualified specialists.

Full-scale measurements of the output voltage on the photodiodes of the colorimeter corresponding to red, blue and green colors were carried out.

Key words: color models; RGB model; CMYK model; digital colorimeter; microcontroller.

1 tabl. 3 fig. Ref: 8 items.

УДК 615.472

Дослідження показників колірних об'єктів за допомогою мікроконтролера STM32F407VG / *В.В. Семенець, О.В. Григор'єв* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 217. С. 128 – 132.

Розглянуто та проаналізовано широко використовувані моделі кольорів, а саме: модель RGB та субтрактивні моделі, зокрема CMY та CMYK. Досліджено характеристики колориметрів, які базуються на використанні моделей RGB, CMY та CMYK. Здійснено порівняльний аналіз переваг і недоліків моделей RGB та CMYK.

Розроблений цифровий колориметр дозволяє здійснювати експрес-контроль кольорових характеристик з високою метрологічною точністю і не потребує високого рівня кваліфікації обслуговуючого персоналу. Зазначений прилад відтворює сприйняття кольору людським зором, і його використання призначено як для осіб з нормальним, так і з порушенням зору. Великою перевагою цього пристрою є його мобільність, швидкість вимірювань і відсутність необхідності у спеціалізованій лабораторії та висококваліфікованих фахівцях.

Проведено повномасштабні вимірювання вихідної напруги на фотодіодах колориметра, що відповідають червоному, синьому та зеленому кольорам.

Ключові слова: моделі кольору; модель RGB; модель CMYK; цифровий колориметр; мікроконтролер.

Табл. 1. Іл. 3. Бібліогр.: 8 назв.

UDC 621.396

About the possibility of protecting UAVs from suppression of control signals / *I.M. Mytsenko, Yu.A. Pedenko, A.N. Roenko* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №217. P. 133 – 138.

Unmanned aerial vehicles (UAVs) are widely used in a wide variety of sectors, both in the national economy and for military purposes. In the agricultural industry, they are used to monitor farmland, analyze soil samples, and even for livestock grazing. In search and rescue operations, UAVs are used to rescue people from life-threatening situations, and each time this technology proves its usefulness more than ever. UAVs are gradually finding their application in retail trade, transportation, entertainment, home security and even in construction using 3D printers. Simultaneously with the growth of the capabilities of UAVs, methods and means were developed to counteract them. The most affordable method of combating an UAV is to suppress the radio signals used to control it. As a result, the UAV loses contact with the operator, which leads to malfunctions in its operation. Therefore, the relevance of developing methods and means of combating suppression of control signals cannot be doubted and is becoming more and more necessary. The purpose of the work is to develop a scheme of the device (protection unit) that protects the UAV from interference radio signals in the operating mode.

The peculiarity of the developed device is that in the absence of interference signals on the standard frequency of the UAV, the protection unit receives the operator's control signal and transmits it to the input of the UAV receiver, which moves in the operating mode and performs the tasks assigned to it. In the event of the appearance of interference radio signals on the standard frequency, it independently detects them and, using the UAV transmitter, issues a command to the operator to change the operating frequency to an additional frequency. At the same time, the output frequency of the protection unit remains constant and is equal to the standard frequency of the UAV. Thus, the protection unit is a simple and inexpensive device that detects interference signals and automatically switches the control of the UAV to an additional operating frequency. A functional diagram of the receiving device was developed, and a detailed description of its operation was provided.

Key words: unmanned aerial vehicle; radio signal suppression; protection unit; communication channel.

3 fig. Ref: 15 items.

УДК 621.396

Про можливість захисту БПЛА від придушення сигналів управління / *І.М. Мищенко, Ю.О. Педенко, О.М. Роєнко* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 217. С. 133 – 138.

Безпілотні літальні апарати (БПЛА) широко застосовують у найрізноманітніших секторах як народного господарства так і у військових цілях. В сільськогосподарській промисловості вони використовуються для моніторингу сільськогосподарських угідь, аналізу зразків ґрунту та навіть для пасіння худоби. В пошуково-рятувальних роботах БПЛА використовуються для рятування людей із ситуацій, що загрожують життю, і щоразу ця технологія доводить свою корисність як ніколи. Поступово БПЛА знаходять своє застосування в роздрібній торгівлі, транспорті, розвагах, охороні житла та навіть у будівництві з використанням 3D-принтерів. Одночасно зі зростанням можливостей БПЛА, розвивалися методи і засоби, що їм протидіють. Найдоступнішим методом боротьби з БПЛА є придушення радіосигналів, які використовуються для його управління. У результа-

ті БПЛА втрачає зв'язок з оператором, що призводить до порушень його функціювання. Тому актуальність розробки методів і засобів боротьби з придушенням сигналів керування не підлягає сумніву та стає все більш необхідною. Метою роботи є розробка схеми пристрою (блоку захисту), що здійснює захист БПЛА від радіосигналів завад у робочому режимі.

Особливість розробленого пристрою полягає в тому, що за відсутністю сигналів завад на штатній частоті БПЛА блок захисту, приймає сигнал керування оператора і передає його на вхід приймача БПЛА, який рухається в робочому режимі і виконує поставлені перед ним завдання. У разі появи радіосигналів завад на штатній частоті, самостійно їх виявляє і за допомогою передавача БПЛА видає команду оператору на зміну робочої частоти на додаткову частоту. При цьому вихідна частота блоку захисту залишається постійною і дорівнює штатної частоті БПЛА. Таким чином, блок захисту являє собою простий і недорогий пристрій, який виявляє сигнали завади й автоматично перемикав керування БПЛА на додаткову робочу частоту Розроблено функціональну схему приймального пристрою, надано детальний опис його роботи.

Ключові слова: безпілотний літальний апарат; придушення радіосигналу; блок захисту; канал зв'язку.

Іл. 3. Бібліогр.: 15 назв.

PHYSICS OF DEVICES, ELEMENTS AND SYSTEMS
ФІЗИКА ПРИЛАДІВ, ЕЛЕМЕНТІВ І СИСТЕМ

UDC 621.357

Study of thermal properties of electronic modules on combined boards with polyimide dielectrics / *V.M. Borshchov, O.M. Listratenko, M.I. Slipchenko, M.A. Protsenko, I.T. Tymchuk, O.V. Kravchenko, I.V. Borshchov //* Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №217. P. 139 – 147.

This article presesnts the construction and theoretical research of thermal models of electronic modules with increased power based on combined boards on aluminum bases using serial heat-welding polyimide - fluoroplastic films, including films with thermal conductivity from 0.12 to 0.46 W/m•K, as well as with the lacquer foil dielectrics with thermal conductivity of PI layers of the order of 4.0 - 4.5 W/(m•K) improved by the authors

Designs and quality test structures of electronic modules were developed. Experimental studies of the efficiency of heat removal from semiconductor devices in quality test structures based on various types of combined boards with polyimide dielectrics were performed.

Technical solutions of combined boards based on multi-layer heat-conductive heat-welding PMF film Kapton®120FMT616 30 μm thick with fluoropolymer double-sided coatings with a thermal conductivity of 0.46 W/(m•K) and combined circuit boards based on improved one-sided lacquer foil copper - polyimide dielectrics with a thickness of highly thermally conductive composite PI layers 60 μm with a thermal conductivity of up to 4.0 - 4.5 W/(m•K), provided under favorable operating conditions with natural unobstructed convection and temperature environment Ta = 25°C the best thermal characteristics of electronic modules from the point of view of maintaining recommended operating temperatures < 70 – 80 °C for high reliability of operation and increased service life.

Key words: thermal models; combined boards on aluminum bases; polyimide composites; quality test structures.

2 tabl. 4 fig. Ref: 11 items.

УДК 621.357

Дослідження теплових властивостей електронних модулів на комбінованих платах з поліімідними діелектриками / *В.М. Борщов, О.М. Лістратенко, М.І. Сліпченко, М.А. Проценко, І.Т. Тимчук, О.В. Кравченко, І.В. Борщов* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 217. С. 139 – 147.

Виконано побудову та теоретичні дослідження теплових моделей електронних модулів з підвищеною потужністю на основі комбінованих плат на алюмінієвих основах з використанням серійних термозварювальних поліімід-фторопластовими плівок, у тому числі з теплопровідністю плівок від 0,12 до 0,46 Вт/м•К, а також з удосконаленими авторами лакофольговими діелектриками з теплопровідністю ПІ шарів порядку 4,0 – 4,5 Вт/(м•К).

Розроблено конструкції та виготовлені тестові структури якості електронних модулів. Виконано експериментальні дослідження ефективності відводу тепла від напівпровідникових пристроїв у тестових структурах якості на основі різних типів комбінованих плат з поліімідними діелектриками.

Технічні рішення комбінованих плат на основі багатошарової теплопровідної термозварювальної ПМФ плівки Kapton®120FMT616 товщиною 30 мкм з фторполімерними двосторонніми покриттями з теплопровідністю 0,46 Вт/(м·К) та комбінованих плат на основі удосконалених односторонніх лакофольгових мідь-поліімідних діелектриків з товщиною високотеплопровідних композиційних ПІ шарів 60 мкм з теплопровідністю до 4,0 – 4,5 Вт/(м•К), забезпечили за сприятливих умов експлуатації при природній неутрудненій конвекції та температурі навколишнього середовища Та = 25°С найкращі теплові характеристики електронних модулів з точки зору рекомендованих робочих температур < 70 – 80 °С для підтримки їх високої надійності роботи та підвищення строків експлуатації.

Ключові слова: теплові моделі; комбіновані плати на алюмінієвих основах; поліімідні композити; тестові структури якості.

Табл. 2. Іл. 4. Бібліогр.: 11 назв.

RADIO ELECTRONIC SYSTEMS
РАДІОЕЛЕКТРОННІ СИСТЕМИ

UDC 621.383

Experimental studies of the characteristics of a resonant leader emitter with a single-pass amplifier / *A.A. Zarudnyi //* Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №217. P. 148 – 153.

The results of experimental studies of the energy and spatial characteristics of the lidar transmitter, built according to the generator-amplifier scheme based on an organic dye with tube pumping, are presented. The choice of a single-pass traveling wave amplifier used in the experiments was determined by preserving the spectral purity of the radiation. When constructing a lidar emitter according to the generator-amplifier scheme under conditions of constant pumping density, the problem arises of choosing the ratio between the length of the active element of the generator and the length of the active medium of the traveling wave amplifier, which would ensure the maximum efficiency of the entire emitter. The main task of the work was the experimental verification of the results of theoretical studies of the narrow-band emitter of the resonant lidar in order to determine the factors affecting the choice of the ratio of the lengths of the active elements of the generator and the amplifier based on the organic dye rhodamine 6Zh with tube pumping with limited total length. The amount of effective radiated energy was used as the main criterion for evaluating the efficiency of the generator-amplifier system.

The experimental results confirm the theoretical conclusions that there are optimal ratios of the lengths of the generator and the amplifier, at which the effective radiation energy is maximal. The limiting length of the amplifier and the energy of the emitter built according to the generator-amplifier circuit are limited by the increase in the intensity of the radiation amplified along the active element, as well as the increase in the intensity of the amplified noise.

Key words: lidar; oscillator; amplifier; flashlamp; resonator; pump.

4 fig. Ref: 15 items.

УДК 621.383

Експериментальні дослідження характеристик випромінювача резонансного лідару з однопроходовим підсилювачем / *О.А. Зарудний* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 217. С. 148 – 153.

Наведено результати експериментальних досліджень енергетичних та просторових характеристик передавача лідару, побудованого за схемою генератор-підсилювач на органічному барвнику з ламповим накачуванням. Вибір однопроходового підсилювача біжучої хвилі, що використовувався в експериментах, був обумовлений зберіганням спектральної чистоти випромінювання. При побудові випромінювача лідара за схемою генератор-підсилювач в умовах постійної щільності накачування виникає проблема вибору співвідношення між протяжністю активного елемента генератора і протяжністю активного середовища підсилювача біжучої хвилі, яка б забезпечувала максимальний ККД всього випромінювача. Основне завдання роботи - експериментальна перевірка результатів теоретичних досліджень вузькосмугового випромінювача резонансного лідару з метою визначення факторів, що впливають на вибір співвідношень довжин активних елементів генератора та підсилювача на основі органічного барвника родамін 6Ж з ламповим накачуванням при обмеженій їх сумарній протяжності. Як основний критерій оцінки ефективності системи генератор-підсилювач використовувалася величина ефективної випромінюваної енергії.

Результати експериментів підтверджують теоретичні висновки про те, що існують оптимальні співвідношення довжин генератора і підсилювача, при яких ефективна енергія випромінювання є максимальною. Гранична довжина підсилювача та енергія випромінювача, побудованого за схемою генератор-підсилювач обмежуються за рахунок зростання інтенсивності випромінювання, що посилюється уздовж активного елементу, а також збільшення інтенсивності посиленого шуму.

Ключові слова: лідар; генератор; підсилювач; лампа-спалах; резонатор; накачка.

Іл. 4. Бібліогр.: 15 назв.

UDC 621.396.96

Method for adapting radioacoustic sounding systems of the atmosphere / *A.V. Kartashov, I.E. Kondrashov* // Radiotekhnika : All-Ukr. Sci. Interdep. Mag. 2024. №217. P. 154 – 163.

Stations of radioacoustic sounding (RAS) of the atmosphere are a promising means of obtaining information about the altitudinal distribution of meteorological parameters in the Earth's atmosphere used in the process of solving current scientific and applied tasks to ensure aircraft flights, weather forecasting, etc. However, the effectiveness of the existing radio-acoustic means is insufficient, and there are practical needs for the development of appropriate prospective structures and algorithms, which will be implemented during the construction of specific stations designed to solve actual applied tasks.

The article presents the synthesis of the RAS systems algorithm adaptation performed by changing the frequency of the sounding radio signal to fulfill the Bragg condition when the emitted acoustic pulse signal moves along the sounding path from the standpoint of optimal control theory. Since in the problem of synthesizing the algorithm for controlling the frequency of a radio signal, the disturbing and leading to violations of the Bragg condition during the propagation of an acoustic packet along the sounding path, as well as the process causing measurement errors, are considered as random, this problem is considered as a problem of stochastic optimal control.

It is shown that in accordance with the separation theorem, known from the theory of optimal stochastic control, the method of controlling the frequency of a sounding radio signal should include sequentially peformed operations of forming estimates of the information parameter of the scattered radio signal, optimal linear filtering of the obtained estimates, and deterministic control of the frequency of the sounding radio signal.

Key words: remote sensing of the atmosphere; method; algorithm; estimation of parameters; management; filtration; temperature; probing signal; synthesis.

4 fig. Ref: 34 items.

УДК 621.396.96

Метод адаптації систем радіоакустичного зондування атмосфери / *О.В. Карташов, І.Є. Кондрашов* // Радіотехніка : Всеукр. міжвід. наук.-техн. зб. 2024. Вип. 217. С. 154 – 163

Системи радіоакустичного зондування (РАЗ) атмосфери використовуються для отримання інформації про висотний розподіл метеопараметрів в атмосфері Землі, яка використовується в процесі вирішення актуальних науково-прикладних завдань з забезпечення польотів літальних апаратів, прогнозу погоди та ін. Проте ефективність існуючих радіоакустичних засобів є недостатньою, і існують потреби практики з розробці відповідних перспективних структур та алгоритмів, що реалізуватимуться при побудові конкретних станцій, призначених для вирішення актуальних прикладних завдань. Основним серед існуючих обмежень систем радіоакустичного зондування є порушення умов Брегга по трасі зондування.

Виконано синтез методу адаптації систем РАЗ шляхом зміни частоти зондувального радіосигналу з метою виконання умови Брегга у міру переміщення випромінюваного акустичного імпульсного сигналу трасою зондування з позицій теорії оптимального управління. Так як у задачі синтезу алгоритму управління частотою радіосигналу процес, що обурює та призводить до порушень умови Брегга при поширенні акустичного пакета по трасі зондування, а також процес, що викликає похибки вимірювання, розглядаються як випадкові, то дану задачу розглянуто як завдання стохастичного оптимального управління.

Показано, що у відповідності з теоремою розділення, відомої з теорії оптимального стохастичного управління, метод управління частотою зондувального радіосигналу повинен включати послідовно виконувані операції формування оцінок інформаційного параметра розсіяного радіосигналу, оптимальної лінійної фільтрації отриманих оцінок і детермінованого управління частотою зондуючого радіосигналу.

Ключові слова: дистанційне зондування атмосфери; метод; алгоритм; оцінка параметрів; управління; фільтрація; температура; зондуючий сигнал; синтез.

Іл. 4. Бібліогр.: 34 назв.

COLLECTION OF SCIENTIFIC PAPERS
RADIOTEKHNIKA
Issue 217
In English and Ukrainian

ЗБІРНИК НАУКОВИХ ПРАЦЬ
РАДІОТЕХНІКА
Випуск 217
Англійською та українською мовами

Коректор Л.І. Сащенко

.

Підп. до друку 15.06.2024. Формат 60x90/8. Папір офсет. Гарнітура Таймс. Друк. ризограф.
Ум. друк. арк. 10,4. Обл.-вид. арк. 9,9. Тираж 300 прим. Зам. № 123. Ціна договір.

Харківський національний університет радіоелектроніки (ХНУРЕ)
Просп. Науки, 14, Харків, 61166.

Оригінал-макет підготовлено і збірник надруковано у ПФ „Колегіум”,
Свідоцтво про внесення суб’єкта видавничої діяльності до Державного реєстру видавців.
Сер. ДК №1722 від 23.03.2004.