

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ,
МОЛОДЕЖИ И СПОРТА УКРАИНЫ

ХАРЬКОВСКИЙ НАЦИОНАЛЬНЫЙ
УНИВЕРСИТЕТ РАДИОЭЛЕКТРОНИКИ

ISSN 0135-1710

АВТОМАТИЗИРОВАННЫЕ СИСТЕМЫ УПРАВЛЕНИЯ И ПРИБОРЫ АВТОМАТИКИ

**Всеукраинский межведомственный
научно-технический сборник**

Основан в 1965 г.

Выпуск 155

Харьков
2011

СОДЕРЖАНИЕ

ИВ ТИЕКУРА (YVES TIECOURA), ХАХАНОВ В.И., ЧУМАЧЕНКО С.В., ЛИТВИНОВА Е.И. АНАЛИЗ КИБЕРПРОСТРАНСТВА И ДИАГНОСТИРОВАНИЕ ФУНКЦИОНАЛЬНЫХ МОДУЛЕЙ.....	4
КУЛИК И.А., СКОРДИНА Е.М., КОСТЕЛЬ С.В. ГЕНЕРИРОВАНИЕ КОДОВ-СОЧЕТАНИЙ ДЛЯ РЕШЕНИЯ ИНФОРМАЦИОННЫХ ЗАДАЧ И УС.....	15
ДОВБИШ А.С., БАРИЛО О.Б. РОЗПАРАЛЕЛЕНИЙ АЛГОРИТМ ОПТИМІЗАЦІЇ ПАРАМЕТРІВ НАВЧАННЯ СППР ДЛЯ КЕРУВАННЯ СЛАБОФОРМАЛІЗОВАНИМ ТЕХНОЛОГІЧНИМ ПРОЦЕСОМ.....	23
ОКСАНИЧ А.П., ПРИТЧИН С.Э., СЕДИН Е.А. РАЗРАБОТКА МЕТОДИК И УСТРОЙСТВ КОНТРОЛЯ СТРУКТУРНОГО СОВЕРШЕНСТВА КРЕМНИЕВЫХ СТРУКТУР.....	29
ЗАГИРНЯК М.В., ОКСАНИЧ А.П., ПЕТРЕНКО В.Р., ПРИТЧИН С.Э., ТЕРБАН В.А. РАЗРАБОТКА МАТЕМАТИЧЕСКОЙ МОДЕЛИ И ПРОГНОЗНОГО РЕГУЛЯТОРА ДЛЯ ПРОЦЕССА ВЫРАЩИВАНИЯ МОНОКРИСТАЛЛОВ ПОЛУИЗОЛИРУЮЩЕГО АРСЕНИДА ГАЛЛИЯ.....	33
БАРАННИК В.В., КАЛЬЧЕНКО Д.С. МЕТОДОЛОГИЧЕСКИЕ ПРИНЦИПЫ ПРЕДСТАВЛЕНИЯ АПЕРТУР ВО МНОЖЕСТВЕ ОДНОМЕРНЫХ ДВУХОСНОВНЫХ ПОЗИЦИОННЫХ ЧИСЕЛ.....	43
КАНЦЕДАЛ С.А., КОСТИКОВА М.В. ОДИН АЛГОРИТМ РЕШЕНИЯ ЗАДАЧИ О ПОКРЫТИИ.....	49
ПОГРЕБНЯК К.А., КРАВЧЕНКО П.О. МОДЕЛЬ КОМБІНОВАНОЇ ІНФРАСТРУКТУРИ ВІДКРИТИХ КЛЮЧІВ.....	53
ЛОПУХИН Ю.В., БОЧАРОВ Е.В., ВАРЕЦА В.В., ПОЛЕЖАЕВ К.В., МАКАРЕНКО А.Н. СТАТИСТИЧЕСКИЙ АНАЛИЗ ПРЕОБРАЗОВАНИЙ ЧИСЕЛ В ПРЕОБРАЗОВАТЕЛЯХ КОДОВ ПАРАЛЛЕЛЬНОГО ТИПА.....	61
КОВАЛЕВ А.И. ПОСТРОЕНИЕ АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ УПРАВЛЕНИЯ ПРОИЗВОДСТВЕННЫМИ ЗАКАЗАМИ НА ОСНОВЕ АРХИТЕКТУРЫ ОПЕРАЦИОННЫХ СИСТЕМ. ЧАСТЬ 2.....	67
ЛЕВИКІН В.М., КОСТЕНКО О.П., ЗІНЧЕНКО Є.Г. РОЗРОБКА МОДЕЛІ КОМПЛЕКСУ ПРОЕКТНИХ ПАТЕРНІВ МАРКЕТИНГОВИХ ІНФОРМАЦІЙНИХ СИСТЕМ.....	75
УЛЬЯНОВСКАЯ Ю.В. ИССЛЕДОВАНИЕ СВОЙСТВ ОТНОШЕНИЙ НЕЧЕТКОЙ БЛИЗОСТИ ОБЪЕКТОВ В ИНФОРМАЦИОННЫХ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ.....	81
ГАПОН А.И., САВИЦКИЙ С.М. МЕТОД ПРОГРАММНОГО РЕГУЛИРОВАНИЯ ИНЕРЦИОННЫМ ОБЪЕКТОМ ДЛЯ СИСТЕМЫ С ОДНИМ ДАТЧИКОМ И ОДНИМ ИСПОЛНИТЕЛЬНЫМ УСТРОЙСТВОМ.....	87
РЕФЕРАТИ.....	93
ПРАВИЛА ОФОРМЛЕНИЯ РУКОПИСЕЙ ДЛЯ АВТОРОВ НАУЧНО-ТЕХНИЧЕСКОГО СБОРНИКА.....	98

АНАЛИЗ КИБЕРПРОСТРАНСТВА И ДИАГНОСТИРОВАНИЕ ФУНКЦИОНАЛЬНЫХ МОДУЛЕЙ

Описываются инфраструктура и технологии анализа киберпространства, в рамках которого созданы транзакционная граф-модель и метод диагностирования цифровых систем на кристаллах, ориентированные на существенное уменьшение времени поиска дефектов и затрат памяти для хранения матрицы диагностирования путем формирования тернарных отношений в форме тест, монитор, функциональный компонент. Решаются задачи: создание модели цифровой системы в виде транзакционного графа, а также тернарные матрицы активизации функциональных компонентов на тестах относительно выбранного множества мониторов; разработки метода анализа матрицы активизации для поиска дефектов с заданной глубиной и синтеза логических функций для последующей реализации встроенного аппаратного диагностирования дефектов.

1. Анализ киберпространства и актуальные задачи

Чтобы идея материализовалась и завоевала весь мир, она должна быть простой и понятной каждому. Такого рода идея подняла на вершину успеха компании Microsoft, Google, Kaspersky lab, Intel, IBM. Рождается новая персональная или индивидуальная модель киберпространства – виртуальный персональный киберкомпьютер – Virtual Personal Cybercomputer (VPC), родителями которого можно считать с одной стороны реальные нанотехнологии и цифровые системы на кристаллах, а с другой – виртуальные сервисы по хранению, обработке и приему-передаче информации. Мотивация появления данной модели на рынке определяется: 1) Необходимостью создания «индивидуального или личного виртуального компьютера или пространства», которое нельзя потерять или украсть. 2) Нежеланием пользователя и высокой стоимостью дублирования информации и сервисов при наличии у него нескольких гаджетов (планшет, смартфон, ноутбук). 3) Высоким уровнем надежности и информационной безопасности хранения личных данных и сервисов в течение всей жизни пользователя. Перечисленные условия может обеспечить только индивидуальная ячейка в «швейцарском банке», которая должна и будет создаваться в ближайшие три года для каждого человека планеты как Personal Cyberspace Cell (PCC). Две точки зрения в своем развитии сходятся к одному понятию, PCC = VPC, первая исходит со стороны киберпространства (компьютерных наук), вторая – со стороны персонального компьютера (компьютерной инженерии). Практически при полном отсутствии недостатков PCC имеет следующие преимущества: 1) Инвариантность функционирования по отношению к любому аппаратному интерфейсу, соединяющему пользователя с киберпространством. 2) Дружественность и интеллектуальная адаптивность к «хозяину» по формату 24/7 на протяжении всей жизни. 3) Аутентификация пользователя и PCC по отношению к облачным и другим сервисам киберпространства, которая сегодня завязана на конкретную аппаратуру. 4) Надежность и доступность, сохраняемость и безопасность PCC, переносимость и физическая неуязвимость благодаря своей виртуальности. 5) Эффективная реляционная структуризация данных и сервисов с признаками интеллекта для поиска, распознавания и принятия решений. 6) Высокая рыночная привлекательность создания кибербанков и PCC-форматов (шаблонов, стандартов), которые ориентированы и необходимы каждому человеку планеты (в денежном эквиваленте это составляет сотни миллиардов долларов). 7) Возможность создания прототипа виртуального персонального киберкомпьютера ограниченными силами нескольких раскрученных компаний, имеющих выход на World Market, и двумя-тремя университетами. 8) Ориентировочная стоимость таких работ с созданием начальной инфраструктуры банков киберпространства – 0,5 – 1,5 млрд долларов.

Цель – создание индивидуального и виртуального компьютера в киберпространстве для выполнения интеллектуальных транзакций с данными и сервисами, ориентированными на каждого человека.

Задачи: 1) Определение технической инфраструктуры для функционирования виртуального РСС. 2) Создание структурированной базы данных для хранения информации и сервисов. 3) Разработка шаблона РСС в виде набора взаимосвязанных сервисов и инструментов, ориентированных на удовлетворение потребностей пользователя. 4) Разработка системы защиты персонального киберпространства, данных и сервисов, включающей аутентификацию, ключи, цифровую подпись, криптографию. 5) Создание интеллектуальных средств поиска, распознавания и принятия решений в виде совокупности фильтров, ориентированных на конкретного пользователя. 6) Создание прототипа РСС и его тестирование для различных типов пользователей. 7) Предложение прототипа компаниям, которые имеют выход на рынок электронных технологий, а также Public Relations путем выступлений в Internet, TV, на конференциях и семинарах.

Сущность – создание инфраструктуры оптимальной организации индивидуального киберпространства в виде виртуального компьютера, имеющего следующие сервисы: 1) электронная почта и телефония; 2) Internet-браузеры для поиска, распознавания и принятия решений; 3) аудио- и видеопроигрыватели; 4) текстовые и звуковые редакторы; 5) электронный banking и shopping; 6) индивидуальный бизнес-браузер для организации рабочего дня; 7) браузер для организации отдыха, культуры и спорта; 8) traveling-браузер; 9) структурированная реляционная база данных для хранения истории и всех типов данных; 10) внешний интерфейс Public Relations; 11) медицинское обслуживание и сервисы; 12) комплексная система защиты информации и сервисов.

Киберкомпьютер (персональный и виртуальный) – виртуальное отображение в киберпространстве функций персонального компьютера для выполнения интеллектуальных транзакций с данными и сервисами, индивидуально ориентированными на каждого человека. Киберпространство – совокупность взаимодействующих по метрике информационных процессов и явлений, использующих в качестве носителя компьютерные системы и сети. Метрика – способ измерения расстояния в пространстве между компонентами процессов или явлений. Расстояние в киберпространстве – это хог-взаимодействие компонентов процессов или явлений, представленных векторами, которое имеет скалярную проекцию в виде расстояния по Хэммингу. Инвариантами расстояния являются: производная (булева), степень изменения, различия или близости компонентов процесса или явления. Процедуры сравнения, измерения, оценивания, распознавания, тестирования, диагностирования, идентификации и принятия решений сводятся к определению хог-отношения, которое для циклических структур равно нуль-вектору.

Далее предлагается решение практической задачи – диагностирование функциональных блоков цифровой системы на кристалле на основе использования аппарата анализа киберпространства.

2. ТАВ-модель диагностирования дефектных компонентов в SoC

Мотивация определяется рыночной привлекательностью матричного метода поиска дефектов в компонентах (программных и аппаратных) цифровых систем на кристаллах, как самого технологичного, который ориентирован на параллельную обработку данных, что дает возможность существенно уменьшить время диагностического обслуживания при возникновении неисправностей.

Цель исследования – создание модели, метода и их аппаратной реализации, ориентированных на существенное уменьшения времени тестирования и затрат памяти для хранения матрицы диагностирования путем формирования тернарных отношений (тест – монитор – функциональный компонент) внутри одной таблицы или ТАВ: Tests – Assertions – Blocks.

Задачи: 1) Разработка модели цифровой системы в виде транзакционного графа, а также матрицы активизации функциональных компонентов на тестах относительно выбранного множества мониторов [1-6]. 2) Разработка метода анализа матрицы активизации для поиска дефектов с заданной глубиной [4-7]. 3) Синтез логических функций для встроенного диагностирования дефектов [8-11].

Модель тестирования цифровой системы представлена в виде следующего преобразования начального уравнения диагноза, определенного хог-отношением параметров <тест – функциональность – неисправные блоки >:

$$T \oplus F \oplus B = 0 \rightarrow B = T \oplus F \rightarrow B = \{T \times A\} \oplus F \rightarrow B = \{T \times A\} \oplus \{F \times m\},$$

которое оформлено в тернарное матричное отношение компонентов:

$$M = \{\{T \times A\} \times \{B\}\} \leftarrow M_{ij} = (T \times A)_i \oplus B_j.$$

Здесь координата матрицы (таблицы) равна 1, если пара тест-монитор $(T \times A)_i$ проверяет (активизирует) дефекты функционального блока $B_j \in B$.

Модель цифровой системы представлена в виде транзакционного графа

$$G = \langle V, A \rangle, V = \{B_1, B_2, \dots, B_i, \dots, B_n\}, A = \{A_1, A_2, \dots, A_j, \dots, A_m\},$$

где определены множество дуг – функциональных блоков и вершин – мониторов для наблюдения совокупности переменных цифровой системы. Для целей диагностирования на графовую модель накладывается совокупность тестовых сегментов $T = \{T_1, T_2, \dots, T_r, \dots, T_k\}$, которая активизирует транзакционные пути в графе. В общем случае модель тестирования представлена декартовым произведением $M = \langle V \times A \times T \rangle$, она имеет размерность $Q = n \times m \times k$. Чтобы уменьшить объем диагностической информации, предлагается каждому тесту поставить в соответствие монитор, который отвечает за визуализацию пути активизации функциональных блоков, что дает возможность уменьшить размерность модели (матрицы) до $Q = n \times k$ при сохранении всех возможностей отношения триады $M = \langle V \times A \times T \rangle$. Для пары тест-монитор возможны не только взаимно-однозначные соответствия $\langle T_i \rightarrow A_j \rangle$, но и функциональные $\langle \{T_i, T_r\} \rightarrow A_j \rangle$, а также инъективные $\langle T_i \rightarrow \{A_j, A_s\} \rangle$. Такое многообразие соответствий дает возможность дублировать один тестовый сегмент для разных мониторов, равно как и нагружать несколько тестов на один и тот же монитор. При этом ячейка матрицы $M_{ij} = \{0,1\}$ всегда сохраняет свою размерность, равную одному биту.

Аналитическая обобщенная модель матричного диагностирования с использованием механизма мониторов ориентирована на достижение заданной глубины поиска дефектов и представлена в следующем виде:

$$\begin{aligned} M &= f(G, L, T, V, A, t), \\ V &= \{B_1, B_2, \dots, B_i, \dots, B_m\}; \\ L &= \{L_1, L_2, \dots, L_i, \dots, L_n\}; \\ A(t) &= \{A_1, A_2, \dots, A_i, \dots, A_k\}; \\ A &\subseteq L; G = L \times V; k \leq n; \\ T &= \{T_1, T_2, \dots, T_i, \dots, T_p\}. \end{aligned}$$

Здесь V_i – группа операторов кода, нагруженная на вершину L_i (переменная, регистр, счетчик, память) и формирующая ее состояние; G – функциональность, представленная транзакционным графом $G = (L, A) \times V$ в виде декартова произведения множества вершин и дуг; A – совокупность мониторов, как подмножество вершин транзакционного графа $A \subseteq L$. Метод поиска неисправностей функциональных блоков использует предварительно построенную таблицу (матрицу) активизации ТАФБ $M = [M_{ij}]$, где строка есть отношение между тестовым сегментом и подмножеством активизированных блоков

$$T_i \rightarrow A_j \approx (M_{i1}, M_{i2}, \dots, M_{ij}, \dots, M_{in}), M_{ij} = \{0,1\},$$

наблюдаемых на мониторе A_j . Столбец таблицы формирует отношение между функциональным блоком, тестовыми сегментами и мониторами $M_j = V_j \approx f(T, A)$. В механизм мониторов может быть введен параметр модельного времени, который частично усложняет матрицу активизации, указывая временной или модельный такт, на котором выполняет-

ся мониторинг состояния вершины или функционального блока на тест-сегменте $A_j = f(T_i, B_j, t_j)$.

Для диагностирования неисправностей на стадии моделирования определяется обобщенная реакция (вектор-столбец) $m = \{m_1, m_2, \dots, m_i, \dots, m_p\}$ механизма мониторов A на тест-сегменты T путем формирования $m_i = f(T_i, A_j)$. Поиск неисправного ФБ основан на определении хог-операции между вектором состояния ассерций и столбцов таблицы ФН $m \oplus (M_1 \vee M_2 \vee \dots \vee M_j \vee \dots \vee M_n)$. Выбор решения определяется методом хог-анализа столбцов путем выбора совокупности векторов B_j с минимальным числом единичных координат

$$B = \min_{j=1, n} [B_j = \sum_{i=1}^p (B_{ij} \oplus m_i)],$$

формирующих функциональные блоки с неисправностями, проверяемыми на тестовых сегментах. В дополнение к модели матричного диагностирования необходимо описать следующие важные свойства матрицы:

- 1) $M_i = (T_i - A_j)$;
- 2) $\bigvee_{i=1}^m M_{ij} \rightarrow \forall M_j = 1$;
- 3) $M_{ij} \bigoplus_{j=1}^n M_{rj} \neq M_{ij}$;
- 4) $M_{ij} \bigoplus_{i=1}^k M_{ir} \neq M_{ij}$;
- 5) $\log_2 n \leq k \leftrightarrow \log_2 |B| \leq |T|$
- 6) $B_j = f(T, A) \rightarrow B \oplus T \oplus A = 0$.

Свойства означают: 1) Каждая строка матрицы есть соответствие или подмножество декартова произведения (тест-монитор). 2) Дизъюнкция всех строк матрицы дает вектор, равный единицам по всем координатам. 3) Все строки матрицы различны, что исключает тестовую избыточность. 4) Все столбцы матрицы различны, что исключает существование эквивалентных неисправностей. 5) Число строк матрицы должно быть больше двоичного логарифма от числа столбцов, что определяет потенциальную диагностируемость всех блоков. 6) Функция диагностирования блока зависит от совокупного теста и мониторов, которые должны быть минимизированы без нарушения диагнозопригодности.

3. Диагнозопригодность проекта

Что касается качества модели диагностирования функциональных нарушений, то она показывает эффективность использования пары (тест, ассерции) для заданной глубины диагностирования. Оценка качества модели функционально зависит от длины теста $|T|$, числа ассерций наблюдения $|A|$, количества распознаваемых блоков с функциональными нарушениями N_d на общем числе программных блоков N :

$$Q = E \times D = \frac{|\log_2 N|}{|T| \times |A|} \times \frac{N_d}{N}.$$

Эффективность диагностирования есть отношение минимального числа двоичных разрядов, необходимых для идентификации (распознавания) всех блоков, к реальному количеству разрядов кода, представленному произведением длины теста на число ассерций в каждом из них. Если первая дробь оценки равна 1 и все блоки с ФН распознаются ($N_d = N$), то тест и ассерции оптимальны, что доставляет критерию качества модели диагностирования значения, равного 1. Далее рассмотрим пример матрицы ABC-графа,

который представлен на рис. 1 и имеет 14 функциональных блоков – дуг, а также 9 мониторов – вершин.

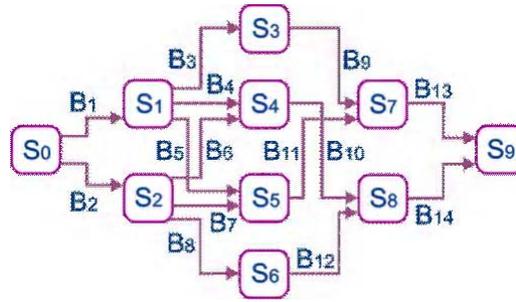


Рис. 1. Пример ABC-графа транзакций

Для такого графа существует 2 решения при создании модели диагностирования дефектов с одной и тремя ассерциями или мониторами:

$$Q_1 = \frac{\lceil \log_2 14 \rceil}{|6| \times |1|} \times \frac{10}{14} = 0,5;$$

$$Q_2 = \frac{\lceil \log_2 14 \rceil}{|6| \times |3|} \times \frac{14}{14} = 0,2.$$

Несмотря на то, что качество модели лучше в первом варианте (из-за меньшего объема таблицы активизации), вторая модель – более предпочтительна, поскольку она имеет максимальную глубину диагностирования, когда все 14 блоков распознаются за счет добавления двух ассерций. Оценка позволяет определить минимальные затраты по длине теста и числу ассерций для создания модели с максимальной глубиной диагностирования.

Интерес представляет оценивание качества структуры кода проекта с позиции диагностируемости (diagnosability) блоков программного продукта. Цель анализа – определить количественную оценку структуры графа и места (вершины) для установки ассерционных мониторов, создающих максимальную глубину диагностирования функциональных нарушений программных блоков. Здесь важна не управляемость и наблюдаемость, как в тестопригодности (testability), а различимость программных блоков с функциональными нарушениями, в пределе представляющая ноль блоков с эквивалентными (неразличимыми) нарушениями. Такая оценка может быть полезной для сравнения графов, реализующих одинаковую функциональность. Здесь необходимо оценивать структуру графа с позиции потенциально заложенной в нем глубины поиска функциональных нарушений программного продукта. Возможным вариантом может быть диагнозопригодность ABC-графа как функция, зависящая от таких смежных дуг при каждой вершине (формирующих число N_n), одна из которых – входящая, другая – исходящая. Такие дуги составляют пути без схождений и разветвлений (N – общее количество дуг в графе):

$$D = \frac{N - N_n}{N}.$$

Каждая вершина, объединяющая 2 дуги, которые входят в число N_n , называется транзитной. Оценка N_n есть число неразличимых функциональных нарушений программных блоков. Места потенциальной установки мониторов для различения ФН – транзитные вершины. С учетом приведенной оценки диагнозопригодности D качество модели диагностирования программного продукта принимает вид:

$$Q = E \times D = \frac{\lceil \log_2 N \rceil}{|T| \times |A|} \times \frac{N - N_n}{N}.$$

Правила синтеза диагнозопригодных программных продуктов: 1) Тест (testbench) должен создавать минимальное число одномерных путей активизации, покрывающих все вершины и дуги ABC-графа. 2) Базовое число мониторов-ассерций равно количеству конечных вершин графа, не имеющих исходящих дуг. 3) В каждой вершине, имеющей одну

входную и одну выходную дугу, может быть размещен дополнительный монитор. 4) Параллельно-независимые блоки кода имеют n мониторов и один тест или один объединенный монитор и n тестов. 5) Последовательно соединенные блоки имеют 1 тест активизации последовательного пути и $n-1$ монитор или n тестов и n мониторов. 6) Вершины графа, имеющие различное число входных и выходных дуг, создают условия для диагностируемости данного участка благодаря одномерным тестам активизации без установки дополнительных мониторов. 7) Совокупность тестовых сегментов (testbench) должна составлять 100%-ное покрытие функциональных режимов (functional coverage), заданных вершинами ABC-графа. 8) Функция диагнозопригодности прямо пропорциональна длине теста, числу ассерций и обратно пропорциональна двоичному логарифму от числа программных блоков:

$$D = \frac{N - N_n}{N} = f(T, A, N) = \frac{|T| \times |A|}{\lceil \log_2 N \rceil}$$

Диагнозопригодность как функция, зависящая от структуры графа (программного продукта), теста и ассерционных мониторов всегда может быть приведена к единичному значению. Для этого существует два альтернативных пути. Первый – увеличение тестовых сегментов, активизирующих новые пути, для различения эквивалентных неисправностей без наращивания ассерций, если структура графа программных блоков имеет такой потенциал связей. Второй – размещение дополнительных ассерционных мониторов в транзитных вершинах графа. Возможен и третий, гибридный вариант, основанный на совместном применении двух перечисленных выше путей. Отношение трех компонентов (число программных блоков, мощность механизма ассерций и длина теста) при единичном значении качества модели диагностирования и диагнозопригодности формирует плоскость оптимальных решений $D = 1 \rightarrow \frac{|T| \times |A|}{\lceil \log_2 N \rceil} = 1 \rightarrow \lceil \log_2 N \rceil = |T| \times |A|$. Она может быть полезной для выбора квазиоптимального варианта альтернативного пути достижения полной различимости на паре $|T| \times |A|$ функциональных нарушений программных блоков.

4. Мультиуровневая модель и метод (движок) диагностирования цифровой системы

Представлена мультидеревом B , где каждая вершина есть трехмерная таблица активизации функциональных модулей, а дуги, исходящие из нее, есть переходы на нижний уровень детализации при диагностировании, когда замена рассматриваемого неисправного функционального блока слишком дорога:

$$B = [B_{ij}^{rs}], \text{ card}B = \sum_{r=1}^n \sum_{s=1}^{m_r} \sum_{j=1}^{k_{rs}} B_{ij}^{rs},$$

n – число уровней мультидерева диагностирования; m_r – количество функциональных блоков или компонентов на уровне r ; k_{rs} – число компонентов в таблице B^{rs} ; $B_{ij}^{rs} = \{0,1\}$ – компонент таблицы активизации, определенный сигналами проверки (непроверки) функционального модуля тест-сегментом T_{i-A_i} относительно наблюдаемого монитора A_i . Каждая вершина-таблица имеет число исходящих вниз дуг, равное количеству функциональных блоков, диагностируемых (представленных) в таблице активизации. Структура мультидерева, соответствующая модели многоуровневого диагностирования, представлена на рис. 2.

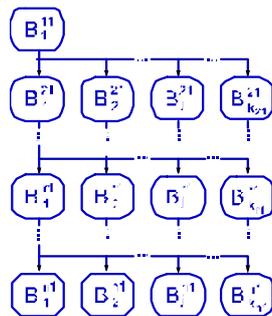


Рис. 2. Фрагмент мультидерева диагностирования цифровых систем

Процесс-модель или метод поиска дефектов по мультидереву диагностирования сводится к созданию движка (рис. 3) для спуска по одной из ветвей дерева на такую глубину, которая удовлетворяет пользователя по степени детализации:

$$B_j^{rs} \oplus A^{rs} = \begin{cases} 0 \rightarrow \{B_j^{r+1,s}, R\}; \\ 1 \rightarrow \{B_{j+1}^{rs}, T\}. \end{cases}$$

Здесь выполняется векторная хог-операция между столбцами матрицы и вектором экспериментальной проверки A^{rs} , который определяется реакцией функциональности, снятой с мониторов (ассерции или разряды регистра граничного сканирования) при подаче всех тест-сегментов. Если хотя бы одна координата полученной векторной хог-суммы равна нулю $B_j^{rs} \oplus A^{rs} = 0$, то выполняется одно из действий: переход к матрице активизации нижнего уровня $B_j^{r+1,s}$ или восстановление работоспособности функционального блока

B_j^{rs} . При этом анализируется, что важнее: 1) время – тогда выполняется ремонт рассматриваемого блока с неисправностью; 2) деньги – тогда осуществляется переход вниз, для уточнения места дефекта, поскольку замена более мелкого блока существенно уменьшает стоимость ремонта. Если хотя бы одна координата полученного вектора хог-суммы равна единице $B_j^{rs} \oplus A^{rs} = 1$, то выполняется переход к анализу следующего столбца матрицы. При нулевых значениях всех координат вектора (ассерционных) мониторов $A^{rs} = 0$ фиксируется исправное состояние всего изделия. Если в рассматриваемой таблице зафиксированы все векторные хог-суммы, не равные нулю $B_j^{rs} \oplus A^{rs} = 1$, то коррекции подлежит тест, построенный для проверки данной функциональности.

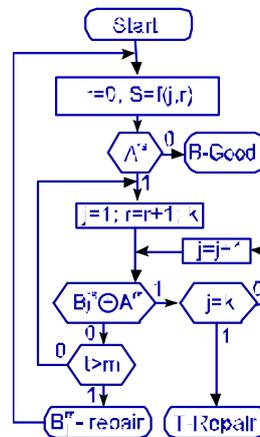


Рис. 3. Движок обхода мультидерева диагностирования

Таким образом, представленная на рис. граф-схема дает возможность эффективно осуществлять сервисное обслуживание сколь угодно сложной технической системы. Преимущества такого движка, инвариантного к уровням иерархии, заключаются в простоте подготовки и представления диагностической информации в виде минимизированной таблицы активизации функциональных блоков на тестовых сегментах.

5. Верификация моделей и метода диагностирования

Для пояснения работоспособности модели и метода далее рассмотрим функциональности трех модулей, входящих в состав цифрового фильтра Добеши [11]. Первым является компонент Row_buffer, для которого создан транзакционный граф на основе RTL-модели (рис. 4). Вершины представлены состояниями переменных и мониторов, отвечающих за входящие в вершину транзакции или дуги, которым соответствуют функциональные блоки.

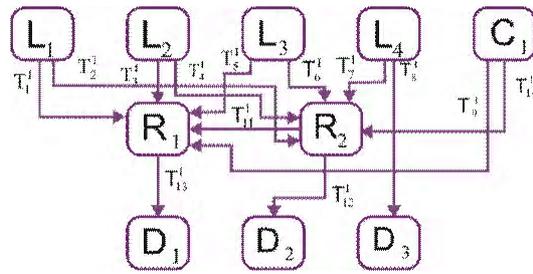


Рис. 4. Компонент Row_buffer транзакционного графа

На основе графа, полученного в процессе моделирования, строится таблица активизации функциональных блоков, строки которой представляют пути активизации блоков к заказанной вершине-монитору. Таблица представляет собой покрытие строками-путями всех столбцов или функциональных блоков. При этом в ней не должно быть хотя бы двух одинаковых столбцов. Отличие таблицы заключается в формировании пары <тест – наблюдаемая вершина>, что дает возможность существенно сократить размерность таблицы при 100% распознавании всех дефектных блоков. Здесь самое главное отличие предложенной модели заключается в возможности описания с помощью таблицы следующих отношений: различные тесты – одна вершина; один тест – различные вершины:

A _{ij}	T ₁	T ₂	T ₃	T ₄	T ₅	T ₆	T ₇	T ₈	T ₉	T ₁₀	T ₁₁	T ₁₂	T ₁₃
t ₁ → D ₃	1
t ₂ → D ₁	1	1
t ₃ → D ₁	.	.	1	1
t ₄ → D ₁	1	1
t ₅ → D ₁	1	.	1
t ₆ → D ₁	1	.	.	.	1
t ₇ → D ₂	.	1	1	.
t ₈ → D ₂	.	.	.	1	1	.
t ₉ → D ₂	1	1	.
t ₁₀ → D ₂	1	1	.
t ₁₁ → D ₂	1	.	.	1	.

С помощью матрицы активизации функциональных блоков (транзакционного графа) и хог-метода поиска дефектов достаточно просто синтезировать логические функции для формирования комбинационной схемы, определяющей в процессе и по результатам моделирования номер функционального блока, который имеет семантические ошибки:

$$D_3 = T_8^1;$$

$$D_1 = T_{13}^1 T_1^1 \vee T_{13}^1 T_3^1 \vee T_{13}^1 T_5^1 \vee T_{13}^1 T_{11}^1 \vee T_{13}^1 T_9^1;$$

$$D_2 = T_{12}^1 T_2^1 \vee T_{12}^1 T_4^1 \vee T_{12}^1 T_6^1 \vee T_{12}^1 T_7^1 \vee T_{12}^1 T_{10}^1.$$

Такое свойство становится возможным благодаря отсутствию эквивалентных неисправностей или одинаковых столбцов в матрице активизации. Поэтому фиксация фактического состояния всех мониторов в вершинах D₁, D₂, D₃ на 11 тестовых наборах дает возможность однозначно идентифицировать некорректный функциональный модуль путем выполнения хог-операции между вектором ассерций и столбцами матрицы активизации. Нулевое значение всех координат результата хог-операции определяет номер столбца, соответствующего неисправному модулю. Имплементация модели и метода в логическую функцию дает возможность определять неисправный блок еще до завершения диагностического эксперимента, если это возможно. Это означает существенную экономию времени

диагностирования отдельных видов дефектов. Например, тест-монитор $t_1 \rightarrow D_3$ дает возможность идентифицировать уже на первом тесте неисправность блока B_8 .

В качестве второго тестового примера для практического использования разработанной модели активизации и хог-метода поиска дефектов далее предлагается синтез матрицы диагностирования для модуля дискретного косинусного преобразования из Xilinx библиотеки, фрагмент которого представлен листингом 1.

Листинг 1. Фрагмент функционального покрытия

```

c0: coverpoint xin
{
  bins minus_big={{128:235}};
  bins minus_sm={{236:255}};
  bins plus_big={{21:127}};
  bins plus_sm={{1:20}};
  bins zero={{0}};
}
c1: coverpoint det_2d
{
  bins minus_big={{128:235}};
  bins minus_sm={{236:255}};
  bins plus_big={{21:127}};
  bins plus_sm={{1:20}};
  bins zero={{0}};
  bins zero2=(0=>0);
}
endgroup
  
```

Для всех 12 модулей фильтра разработаны транзакционные графы, таблицы активизации и логические функции для тестирования и поиска дефектов дискретного косинусного преобразования. Граф с матрицей активизации и логической функцией (рис. 5), также принадлежащей фильтру, представлены ниже.

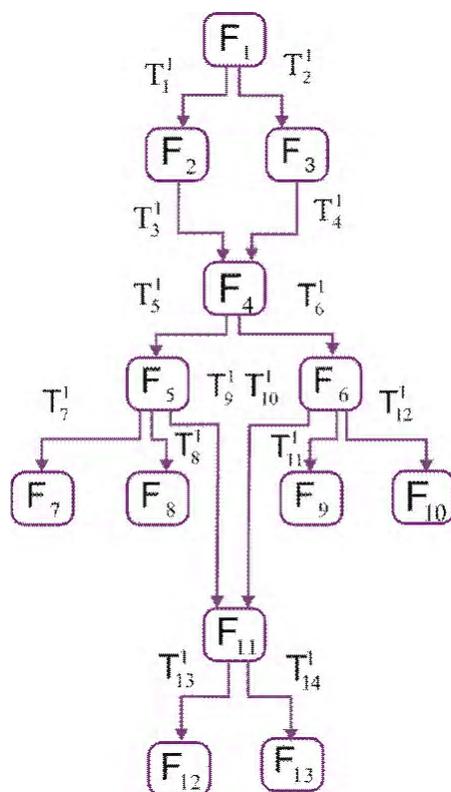


Рис. 5. Транзакционный граф main-RTL

Данному графу ставится в соответствие следующая матрица диагностирования:

A _{ij}	T ₁	T ₂	T ₃	T ₄	T ₅	T ₆	T ₇	T ₈	T ₉	T ₁₀	T ₁₁	T ₁₂	T ₁₃	T ₁₄
P ₁ → F ₇	1	.	1	.	1	.	1
P ₂ → F ₈	.	1	.	1	1	.	.	1
P ₃ → F ₉	1	.	1	.	.	1	1	.	.	.
P ₄ → F ₁₀	.	1	.	1	.	1	1	.	.
P ₅ → F ₁₂	1	.	1	.	1	.	.	.	1	.	.	.	1	.
P ₆ → F ₁₃	.	1	.	1	.	1	.	.	.	1	.	.	.	1
P ₁ → F ₂	1
P ₂ → F ₃	.	1

Система функций диагностирования:

$$F_7 = T_1^1 T_3^1 T_5^1 T_7^1; F_8 = T_2^1 T_4^1 T_5^1 T_8^1; F_9 = T_{11}^1 T_6^1 T_1^1 T_3^1;$$

$$F_{10} = T_4^1 T_5^1 T_6^1 T_{12}^1; F_{12} = T_1^1 T_3^1 T_5^1 T_9^1 T_{13}^1; F_{13} = T_2^1 T_4^1 T_6^1 T_{10}^1 T_{14}^1;$$

$$F_2 = T_1^1; F_3 = T_2^1.$$

Фрагмент механизма мониторов представлен листингом 2.

Листинг 2. Фрагмент кода механизма мониторов

```
sequence first( reg[7:0] a, reg[7:0]b);
reg[7:0] d;
(!RST,d=a)
##7 (b==d);
endsequence
property f(a,b);
@(posedge CLK)
// disable iff(RST||$isunknown(a)) first(a,b);
!RST | => first(a,b);
endproperty
odin:assert property (f(xin,xa7_in))
// $display("Very good");
else $error("The end, xin =%b,xa7_in=%b", $past(xin, 7),xa7_in);
```

В результате тестирования дискретного косинусного преобразования в среде Riviera, Aldec были найдены неточности в семи строках HDL-модели:

```
//add_sub1a <= xa7_reg + xa0_reg;//
```

Последующая коррекция кода привела к листингу 3.

Листинг 3. Исправленный фрагмент кода

```
add_sub1a <= ({xa7_reg[8],xa7_reg} + {xa0_reg[8],xa0_reg});
add_sub2a <= ({xa6_reg[8],xa6_reg} + {xa1_reg[8],xa1_reg});
add_sub3a <= ({xa5_reg[8],xa5_reg} + {xa2_reg[8],xa2_reg});
add_sub4a <= ({xa4_reg[8],xa4_reg} + {xa3_reg[8],xa3_reg});
end
else if (toggleA == 1'b0)
begin
add_sub1a <= ({xa7_reg[8],xa7_reg} - {xa0_reg[8],xa0_reg});
add_sub2a <= ({xa6_reg[8],xa6_reg} - {xa1_reg[8],xa1_reg});
add_sub3a <= ({xa5_reg[8],xa5_reg} - {xa2_reg[8],xa2_reg});
add_sub4a <= ({xa4_reg[8],xa4_reg} - {xa3_reg[8],xa3_reg});
```

6. Заключение

1. Представлены инфраструктура и технологии анализа киберпространства, в рамках которого созданы транзакционная граф-модель и метод диагностирования цифровых сис-

тем на кристаллах, ориентированные на существенное уменьшение времени поиска дефектов и затрат памяти для хранения матрицы диагностирования путем формирования тернарных отношений в форме тест, монитор, функциональный компонент.

2. Предложена усовершенствованная процесс-модель определения функциональных нарушений в программном или аппаратном изделии, которая отличается использованием хог-операции, что дает возможность повысить быстродействие диагностирования одиночных или кратных дефектов (функциональных нарушений) на основе параллельного анализа таблицы неисправностей, стандарта граничного сканирования IEEE 1500 и векторных операций and, or, хог.

3. Представлена модель диагностирования функциональностей цифровой системы на кристалле в форме мультидерева и метод обхода вершин дерева, имплементированный в движок поиска дефектов с заданной глубиной, которая существенно повышает быстродействие сервисного обслуживания программных и аппаратных компонентов промышленных изделий.

4. Выполнена тестовая верификация метода диагностирования на трех реальных примерах, представленных функциональностями фильтра косинусного преобразования цифровой системы на кристалле, которая показала состоятельность полученных результатов для минимизации времени поиска дефектов и памяти для хранения диагностической информации, а также повышения глубины диагностирования цифровых изделий.

Список литературы: 1. *Основы технической диагностики* / Под ред. П.П.Пархоменко. М.: Энергия, 1976. 460с. 2. *Пархоменко П.П., Согомонян Е.С.* Основы технической диагностики (Оптимизация алгоритмов диагностирования, аппаратные средства) / Под ред. П.П. Пархоменко. М.: Энергия. 1981. 320 с. 3. *Инфраструктура мозгоподобных вычислительных процессов* / М.Ф. Бондаренко, О.А. Гузь, В.И. Хаханов, Ю.П. Шабанов-Кушнаренко. Харьков: Новое слово. 2010. 160 с. 4. *Проектирование и верификация цифровых систем на кристаллах* / В.И. Хаханов, И.В. Хаханова, Е.И. Литвинова, О.А. Гузь. Харьков: Новое слово. 2010. 528с. 5. *Семенец В.В., Хаханова И.В., Хаханов В.И.* Проектирование цифровых систем с использованием языка VHDL. Харьков: ХНУРЭ. 2003. 492 с. 6. *Хаханов В.И., Хаханова И.В.* VHDL+Verilog = синтез за минуты. Харьков: ХНУРЭ. 2006. 264с. 7. *IEEE Standard for Reduced-Pin and Enhanced-Functionality Test Access Port and Boundary-Scan Architecture IEEE Std 1149.7-2009.* 985 p. 8. *Da Silva F., McLaurin T., Waayers T.* The Core Test Wrapper Handbook. Rationale and Application of IEEE Std. 1500™. Springer. 2006. XXIX. 276 p. 9. *Marinissen E.J., Yervant Zorian.* Guest Editors' Introduction: The Status of IEEE Std 1500 // IEEE Design & Test of Computers. 2009. No26(1). P.6-7. 10. *Benso A., Di Carlo S., Prinetto P., Zorian Y.* IEEE Standard 1500 Compliance Verification for Embedded Cores // IEEE Trans. VLSI. 2008. No 16(4). P. 397-407. 11. *Хаханов В.И., Литвинова Е.И., Чумаченко С.В. Гузь О.А.* Логический ассоциативный вычислитель. Электронное моделирование. 2011. № 1. С.73-83.

Поступила в редколлегию 02.06.2011

Тиекура Ив (Tiecoura Yves), аспирант кафедры АПВТ ХНУРЭ. Научные интересы: техническая диагностика цифровых систем и сетей. Адрес: Украина, 61166, Харьков, пр. Ленина, 14, тел. 70-21-326. E-mail: hahanov@kture.kharkov.ua.

Хаханов Владимир Иванович, декан факультета КИУ ХНУРЭ, д-р техн. наук, профессор кафедры АПВТ ХНУРЭ. Научные интересы: техническая диагностика цифровых систем, сетей и программных продуктов. Увлечения: баскетбол, футбол, горные лыжи. Адрес: Украина, 61166, Харьков, пр. Ленина, 14, тел. 70-21-326. E-mail: hahanov@kture.kharkov.ua.

Чумаченко Светлана Викторовна, д-р техн. наук, профессор кафедры АПВТ ХНУРЭ. Научные интересы: математическое моделирование, теория рядов, методы дискретной оптимизации. Адрес: Украина, 61166, Харьков, пр. Ленина, 14, тел. 70-21-326. E-mail: ri@kture.kharkov.ua.

Литвинова Евгения Ивановна, д-р техн. наук, профессор кафедры АПВТ ХНУРЭ. Научные интересы: автоматизация диагностирования и встроенный ремонт компонентов цифровых систем в пакете кристаллов. Адрес: Украина, 61166, Харьков, пр. Ленина, 14, тел. 70-21-421. E-mail: kiu@kture.kharkov.ua.

ГЕНЕРИРОВАНИЕ КОДОВ-СОЧЕТАНИЙ ДЛЯ РЕШЕНИЯ ИНФОРМАЦИОННЫХ ЗАДАЧ ИУС

Предлагаются математические модели процессов перечисления и генерирования кодов-сочетаний на основе разработанных биномиальных отображений. Приведено теоретическое обоснование взаимодозначности множеств кодов-сочетаний, биномиальных чисел и соответствующих им номеров, а также операций перехода между ними. Рассматривается связь характеристик предлагаемых моделей перечисления и генерирования кодов-сочетаний с параметрами информационных задач ИУС.

1. Введение

Во многих информационно-управляющих системах (ИУС) достаточно большие объемы информации составляют данные, представленные последовательностями, сумма элементов которых не превосходит или равна некоторой величине, т.е. кодами-сочетаниями. Генерирование таких последовательностей и приписывание им номеров позволяет проводить не только их эффективное кодирование, но и существенно облегчить решение задач целочисленного программирования, в которых генерирование или перечисление допустимого множества векторов является довольно трудной задачей.

Информационные задачи, распространенные в ИУС, в том числе комбинаторных, возможно решать более эффективно, используя коды-сочетания и предлагаемые в этой работе способы их получения. К ним относятся задачи сбора, коммутации, хранения и передачи данных, адаптации параметров и принятия оптимальных управленческих решений. Более того, одни и те же ограничения соответствуют одновременно и монотонным числовым последовательностям, и комбинаторным сочетаниям, тем самым расширяя область возможного эффективного применения кодов-сочетаний на задачи вычислительной обработки данных [1, 2].

Но эффективному применению кодов-сочетаний в ИУС препятствует отсутствие регулярных методов их перечисления и генерирования, которые бы учитывали не только ограничения на элементы рассматриваемых кодов, но и их структурные свойства и числовые характеристики. Последнее особенно важно с точки зрения применения кодов-сочетаний в ИУС, где в процессе контроля и принятия управленческих решений важное место занимают методы вычислительной обработки данных.

В работах [3, 4] отражается общий подход к генерированию комбинаторных конфигураций, к числу которых относятся и сочетания, представленный общими комбинаторными схемами: методом решета и методом поиска с возвращением. Но данные методы не учитывают влияния ограничений на структуру множеств кодов-сочетаний, что приводит к достаточно большим временным и стоимостным затратам при практической их реализации.

В монографии [1] предлагаются системы нумерации и денумерации для сочетаний. В них более полно учитываются специфические ограничения на элементы сочетаний, но без рассмотрения их структурных свойств, без выделения из сочетаний структурных чисел, наделяющих рассматриваемые коды числовыми характеристиками.

Приведенное ниже исследование основывается на установленном в работе [5] положении, что в структуре комбинаторных объектов можно обнаружить соответствующие им структурные числа. В частности, для кодов-сочетаний такими структурными числами являются биномиальные числа.

Цель предлагаемой работы – снижение временных и стоимостных затрат при выполнении характерных для ИУС информационных задач: хранения и передача данных, оптимизация процессов обработки информации и принятие управленческих решений.

Для достижения указанной цели необходимо решить следующие *задачи*:

– выделить в структуре кодов-сочетаний двоичные биномиальные числа;

– разработать математические модели процессов перечисления и генерирования кодов-сочетаний с использованием двоичных биномиальных чисел.

Поставленные задачи следует рассматривать с точки зрения комбинаторного кодирования кодов-сочетаний ввиду следующего обоснования.

1. Коды-сочетания представляют собой последовательности, на значения весовых характеристик и расположение элементов которых накладываются ограничения. Вследствие этого операциям кодирования и декодирования кодов-сочетаний придается комбинаторный характер [1, 6].

2. Комбинаторный подход к представлению кодов-сочетаний подразумевает учет их специфических особенностей, что позволит избежать громоздких в плане вычислительных и временных затрат общих схем для их получения [3, 4].

3. Комбинаторные ИУС достаточно широко распространены и разрабатываемые по сути комбинаторные модели и методы генерирования и перечисления кодов-сочетаний должны органично войти в математическое и алгоритмическое обеспечение таких информационных систем [1, 7].

2. Предварительные сведения

Комбинаторное решение сформулированных в работе задач представляет собой разработку методов комбинаторного кодирования множества Y кодов-сочетаний, имеющих заданное ограничение R_Y . Это означает установление взаимно-однозначного отображения $f: Y \rightarrow F$ между последовательностями $Y_j \in Y$ и элементами множества $F = \{0, 1, \dots, |Y| - 1\}$ целых неотрицательных чисел $F_j \in F$. В рамках комбинаторного кодирования рассматриваются три задачи: подсчета, перечисления и генерирования [1, 8].

В работе [5] предложен универсальный подход для решения такого типа задач на основе специализированных структурных систем счисления, когда под данный тип комбинаторных конфигураций подбирается соответствующая им структурная система счисления и порождаемые ее структурные числа. В качестве структурных чисел для построения моделей процессов генерирования и перечисления кодов-сочетаний предлагается применять биномиальные числа.

Биномиальные числа генерируются биномиальными системами счисления, которые относятся к классу структурных позиционных систем [5, 9]. Особенностью таких систем счисления является то, что в них введены ограничения, которые вес цифр ставят в зависимость не только от их позиции в структурном (биномиальном) числе, но и от значений предшествующих им цифр. В качестве весов цифр в биномиальных числах используются биномиальные коэффициенты.

Двоичные биномиальные числа $X_j = (x_1 x_2 \dots x_i \dots x_r)$ генерируются двоичной биномиальной системой счисления, обладающей числовой функцией вида:

$$F_j = \sum_{i=1}^r x_i C_{n-i}^{k-q_i}, \quad (1)$$

и системами ограничений для образования множества X чисел $X_j, X_j \in X, j = \overline{0, N-1}$:

$$\begin{cases} q = k \\ x_r = 1 \end{cases} \text{ и } \begin{cases} l = n - k \\ x_r = 0 \end{cases}, \quad (2)$$

где n и k – целочисленные параметры двоичной биномиальной системы счисления; r – количество разрядов биномиального числа, $r < n$; x_i – биномиальная двоичная цифра $x_i \in \{0, 1\}$; q – число единиц в биномиальном числе; q_i – сумма единичных x_i , начиная с первого разряда числа до $(i-1)$ -го включительно:

$$q_i = \sum_{t=1}^{i-1} x_t, \quad (3)$$

здесь $q_1 = 0, q_i \leq k$; $N = C_n^k$ – количество двоичных биномиальных чисел X_j .

3. Связь решаемых задач с параметрами ИУС

В рамках временных и стоимостных характеристик реализации задач ИУС, связанных с кодами-сочетаниями, можно выделить составляющие, отвечающие за генерирование, перечисление и обработку рассматриваемых кодов:

$$T = \tau_{\text{КСЧ}} + T_{\text{КСЧ}}, \quad V = v_{\text{КСЧ}} + V_{\text{КСЧ}},$$

где T и V – время и объем аппаратно-программных затрат для решения типовой информационной задачи соответственно; $\tau_{\text{КСЧ}}$ и $v_{\text{КСЧ}}$ – время и объем аппаратно-программных затрат соответственно, требуемых на генерирование, перечисление и обработку кодов-сочетаний; $T_{\text{КСЧ}}$ и $V_{\text{КСЧ}}$ – время и объем аппаратно-программных затрат для решения заданной задачи, но с использованием кодов-сочетаний. Следовательно, в целях увеличения производительности ИУС необходимо и возможно минимизировать переменные $\tau_{\text{КСЧ}}$ и $v_{\text{КСЧ}}$ на основе предлагаемых новых математических моделей и алгоритмов, разрабатываемых на базе комбинаторного подхода к представлению кодов-сочетаний через биномиальные числа:

$$T_{\min} = \min \tau_{\text{КСЧ}} + T_{\text{КСЧ}}, \quad V_{\min} = \min v_{\text{КСЧ}} + V_{\text{КСЧ}}.$$

В качестве примера ИУС, в которой задействованы коды-сочетания, можно привести систему контроля ошибок на основе t -SEC/AUDE кодов, где t – минимальное граничное полурасстояние (half-distance) между двумя произвольными кодовыми комбинациями. Данная система, с одной стороны, обнаруживает все асимметричные ошибки, которые характерны для работы электронных устройств, а с другой – обнаруживает и исправляет все однократные симметричные ошибки, которые возникают в основном при передаче по каналам связи. Для формирования t -SEC/AUDE кодов широко используются коды-сочетания как с одним весом, так и с двумя [10].

Аналогично увеличение производительности при разрабатываемом подходе будет наблюдаться и для ИУС распределения частот передачи в сетях радиосвязи, в том числе мобильной связи стандарта GSM. Для формирования списков распределения частот радиопередатчикам также эффективно можно использовать коды-сочетания [11].

4. Модели процессов генерирования и перечисления кодов-сочетаний

Использование биномиальных чисел в процессах генерирования и перечисления кодов-сочетаний обосновывается тем, что структуры кодов-сочетаний и биномиальных чисел аналогичны, и приемы, применяемые для получения их множеств, являются во многом сходными. Суть предлагаемого подхода заключается в том, что коды-сочетания наделяются числовыми характеристиками, например количественным эквивалентом соответствующих им биномиальных чисел, порождаемых биномиальными системами счисления. В результате возможно не только ускорение процессов генерирования и перечисления кодов-сочетаний, но и расширение возможностей по их преобразованию. В частности, весьма заметно упрощается получение кодов-сочетаний не только в систематическом, но и в случайном порядке.

Для решения задачи по повышению производительности ИУС, применяющих коды-сочетания, предлагается использовать биномиальные отображения.

Определение. Отображение $f: X \rightarrow Y$ называется биномиальным, если его область определения X или область Y значений составляют биномиальные числа.

Определив соответствие $G \subseteq X \times Y$, $(X_j, Y_j) \in G$, между множествами X двоичных биномиальных чисел и Y кодов-сочетаний, тем самым задачи пересчета и генерирования кодов-сочетаний Y_j свдем к аналогичным задачам, но уже для биномиальных чисел X_j . Обратное, построив соответствие $Z \subseteq X \times Y$, $(Y_j, X_j) \in Z$, решение задачи перечисления кодов-сочетаний Y_j будем основывать на результатах решения этой же задачи, но для двоичных биномиальных чисел X_j . С учетом существования числовой функции (1) и систем ограничений (2) для биномиальных чисел X_j , которые являются числообразующими, решение таких задач на основе предложенного подхода потребует заметно меньшего объема вычислительных и временных затрат. Очевидно, с точки зрения однозначности

кодирования и декодирования кодов-сочетаний соответствия G и Z должно быть функциональными и определять прямую φ и обратную φ^{-1} функции перехода от биномиальных чисел X_j к кодам-сочетаниям Y_j и обратно от Y_j к X_j . В свою очередь, нахождение функций $Y_j = \varphi(X_j)$ и $X_j = \varphi^{-1}(Y_j)$ означает построение биективных биномиальных отображений $\varphi: X \rightarrow Y$ и $\varphi^{-1}: Y \rightarrow X$.

Таким образом, математическими моделями исследуемых процессов являются соотношения, представляющие собой следующие сложные функции:

1) перечисление кодов-сочетаний Y_j с заданным ограничением R_Y в целях получения их номеров F_j :

$$F_j = f(Y_j[R_Y]) = \psi(\varphi^{-1}(Y_j[R_Y])); \quad (4)$$

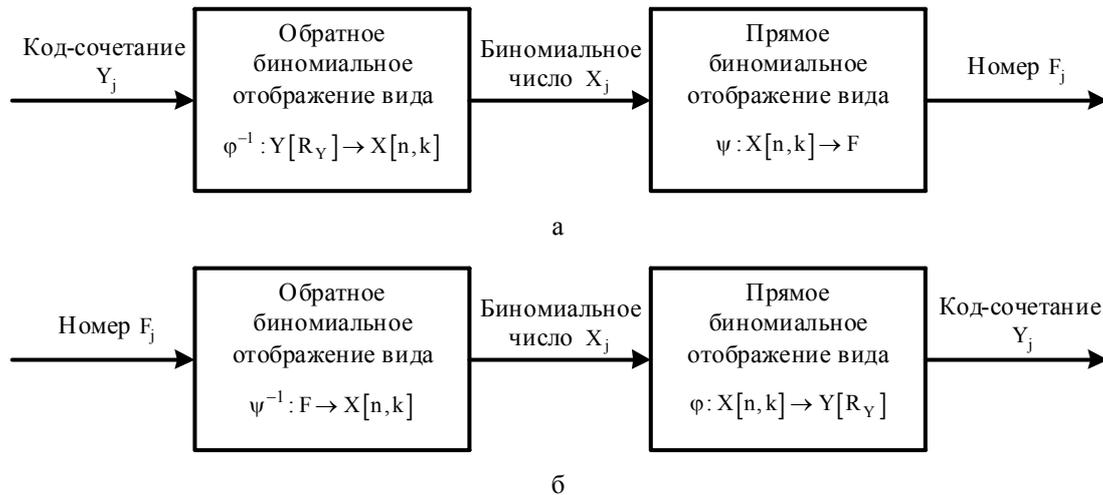
2) генерирование кодов-сочетаний Y_j с заданным ограничением R_Y на основе соответствующих номеров F_j :

$$Y_j[R_Y] = f^{-1}(F_j) = \varphi(\psi^{-1}(F_j)); \quad (5)$$

где $\psi: X \rightarrow F$ и $\psi^{-1}: F \rightarrow X$ – биективные прямое и обратное отображения множества X биномиальных чисел X_j на множество F номеров F_j и обратно F на X соответственно.

Следует отметить, что отображения ψ и ψ^{-1} могут быть реализованы с использованием числовой функции (1) и систем ограничений (2) [5, 9].

В графическом виде модели процессов перечисления и генерирования кодов-сочетаний $Y_j \in Y[R_Y]$ представлены на рисунке.



Модели процессов: а – перечисления; б – генерирования кодов-сочетаний

От задач перечисления и генерирования кодов-сочетаний общего вида перейдем к частному случаю, когда требуется получение отдельного вида сочетаний – квазиравновесных комбинаций $Y_j = (y_1 y_2 \dots y_i \dots y_{n-1})$ с весовым распределением количества единиц $k-1, k$, где $1 \leq k \leq n-1$. При этом ограничение R_Y будет иметь вид вектора, состоящего из двух компонентов $k-1$ и k :

$$R_Y = (k-1, k), \quad k = \sum_{i=1}^{n-1} y_i.$$

Прямое и обратное биномиальные отображения $\varphi: X[n, k] \rightarrow Y[(k-1, k)]$ и $\varphi^{-1}: Y[(k-1, k)] \rightarrow X[n, k]$ для квазиравновесных кодов-сочетаний $Y_j = (y_1 y_2 \dots y_i \dots y_{n-1})$ и двоичных биномиальных чисел $X_j = (x_1 x_2 \dots x_i \dots x_r)$ длины $\min(k, n-k) \leq r \leq n-1$ с пара-

метрами n и k обосновываются теоремами 1 и 2, доказательства которых приведены в работах [12, 13]. В отличие от [12, 13] формулирование данных теорем выполняется в более формализованном виде.

Условимся, что $Y_j^{y_{n-1}=0}$ и $Y_j^{y_{n-1}=1}$ – квазиравновесные комбинации, последние разряды которых $y_{n-1} = 0$ и $y_{n-1} = 1$ соответственно. Подобные обозначения аналогично вводятся и для биномиальных чисел X_j . Кроме того, обозначим операцию конкатенации кодов как “++”, а операцию рассоединения в виде “--”.

Теорема 1. О прямом биномиальном отображении Φ квазиравновесных комбинаций.

$$\text{Если } q = k \text{ и } x_r = 1, \text{ то } Y_j^{y_{n-1}=0} = X_j^{x_r=1} ++ (00\dots 0) = (y_1 y_2 \dots y_i \dots y_{n-2} 0).$$

$$\text{Если } l = n - k \text{ и } x_r = 0, \text{ то } Y_j^{y_{n-1}=1} = X_j^{x_r=0} ++ (11\dots 1) = (y_1 y_2 \dots y_i \dots y_{n-2} 1).$$

В остальных случаях

$$Y_j^{y_{n-1}=0} = X_j^{x_{n-1}=0} = (y_1 y_2 \dots y_i \dots y_{n-2} 0) \text{ и } Y_j^{y_{n-1}=1} = X_j^{x_{n-1}=1} = (y_1 y_2 \dots y_i \dots y_{n-2} 1),$$

где $j = \overline{0, C_n^k - 1}$, а q и l – количество двоичных единиц и нулей в X_j соответственно.

Теорема 2. Об обратном биномиальном отображении Φ^{-1} квазиравновесных комбинаций.

$$\text{Если } q = k \text{ и } y_{n-1} = 0, \text{ то } X_j^{x_r=1} = Y_j^{y_{n-1}=0} -- (00\dots 0) = (x_1 x_2 \dots x_i \dots x_{r-1} 1).$$

$$\text{Если } q = k - 1 \text{ и } y_{n-1} = 1, \text{ то } X_j^{x_r=0} = Y_j^{y_{n-1}=1} -- (11\dots 1) = (x_1 x_2 \dots x_i \dots x_{r-1} 0).$$

В остальных случаях

$$X_j^{x_{n-1}=0} = Y_j^{y_{n-1}=0} = (x_1 x_2 \dots x_i \dots x_{n-2} 0) \text{ и } X_j^{x_{n-1}=1} = Y_j^{y_{n-1}=1} = (x_1 x_2 \dots x_i \dots x_{n-2} 1),$$

где $j = \overline{0, C_n^k - 1}$, а q и l – количества двоичных единиц и нулей в Y_j соответственно.

Пример 1. Пусть 01000 и 1011 двоичные биномиальные числа с параметрами $n = 7$ и $k = 3$. Необходимо определить соответствующие им квазиравновесные комбинации с ограничением $R_Y = (2, 3)$.

Решение. Для 01000 имеем $l = n - k = 7 - 3 = 4$ и $x_{r=5} = 0$. Тогда согласно теореме 1 исходное биномиальное число дополняем справа единицей до общего количества разрядов $n - 1 = 6$: 010001. Для 1011 имеем $q = k = 3$ и $x_{r=4} = 1$. Тогда согласно теореме 1 исходное биномиальное число дополняем справа двумя нулями до общего количества разрядов $n - 1 = 6$: 101100. В результате получаем квазиравновесные комбинации с ограничением $R_Y = (2, 3)$: 010001 и 101100.

Пример 2. Пусть 010101 и 110100 квазиравновесные комбинации с ограничением $R_Y = (2, 3)$. Необходимо определить соответствующие им двоичные биномиальные числа с параметрами $n = 7$ и $k = 3$.

Решение. Для 010101 имеем $q = k = 3$ и $y_{n-1=6} = 1$. Тогда согласно теореме 2 данная исходная комбинация остается без изменений. Для 110100 имеем $q = k = 3$ и $y_{n-1=6} = 0$. Тогда согласно теореме 2 у исходной комбинации отбрасываем серию нулей до появления единицы: 1101. В результате получаем двоичные биномиальные числа с параметрами $n = 7$ и $k = 3$: 010101 и 1101.

Как демонстрируют соотношения (1) и (2), для перечисления и генерирования кодов-сочетаний с ограничением $R_Y = (k - 1, k)$ помимо построенных отображений Φ и Φ^{-1} должны использоваться биективные биномиальные отображения вида $\Psi: X \rightarrow F$ и $\Psi^{-1}: F \rightarrow Y$, которые в свою очередь представляют собой нахождение десятичных (в двоичном представлении) и биномиальных количественных эквивалентов. Чтобы обосно-

вать совместное применение отображений ψ и φ^{-1} для реализации функции $F_j = f(Y_j)$ и отображений φ и ψ^{-1} для реализации функции $Y_j = f^{-1}(F_j)$, необходимо привести и доказать объединяющие теоремы о возможности и однозначности отображения $f: Y \rightarrow F$, т.е. теорему о перечислении, и отображения $f^{-1}: F \rightarrow Y$, т.е. теорему о генерировании.

Теорема 3. О перечислении кодов-сочетаний с ограничением $R_Y = (k-1, k)$ (квазиравновесных кодов).

Всякой последовательности $Y_j \in Y[(k-1, k)]$, $Y_j = (y_1 y_2 \dots y_i \dots y_{n-1})$, $y_i \in \{0, 1\}$, удовлетворяющей ограничению $R_Y = (k-1, k)$, можно поставить в соответствие число $F_j \in F = \{0, 1, 2, \dots\}$ с помощью отображений:

1) отображения $\varphi^{-1}: Y[(k-1, k)] \rightarrow X[n, k]$ последовательности Y_j на биномиальное число $X_j \in X[n, k]$, $X_j = (x_1 x_2 \dots x_i \dots x_r)$, $x_i \in \{0, 1\}$;

2) отображения $\psi: X[n, k] \rightarrow F$ биномиального числа X_j на номер F_j в соответствии с числовой функцией:

$$F_j = \sum_{i=1}^r x_i \rho_i, \quad (6)$$

где $\rho_i = C_{n-i}^{k-q_i}$, $j = \overline{0, C_n^k - 1}$.

Доказательство. Обоснованием $\varphi^{-1}: Y[(k-1, k)] \rightarrow X[n, k]$ есть теорема 2. Для доказательства существования $\psi: X[n, k] \rightarrow F$, реализуемого функцией (6), произведем последовательное деление номера F_j с остатком. На каждом очередном шаге будем делить остаток τ_{i-1} на ρ_i , получая частное δ_i и остаток τ_i :

$F_j = \delta_1 \rho_1 + \tau_1$, $\tau_1 = \delta_2 \rho_2 + \tau_2, \dots$, $\tau_{i-1} = \delta_i \rho_i + \tau_i, \dots$, $\tau_{r-2} = \delta_{r-1} \rho_{r-1} + \tau_{r-1}$, $\tau_{r-1} = \delta_r \rho_r + \tau_r$, где $i = \overline{1, r}$, а $\rho_i = C_{n-i}^{k-q_i}$ представляет собой весовой коэффициент i -го шага деления. Общее количество X_j с параметрами n и k есть $N = C_n^k$, т.е. значение числа F_j лежит в диапазоне $0 \leq F_j < C_n^k$. Тогда, если $F_j = \delta_1 \rho_1 + \tau_1$, то при $i=1$ получаем $\rho_1 = C_{n-1}^{k-q_1}$ и $F_j = \delta_1 C_{n-1}^{k-q_1} + \tau_1$. Но известным является выражение $C_n^k = C_{n-1}^k + C_{n-1}^{k-1}$. Таким образом, с одной стороны

$$C_n^k = C_{n-1}^k + C_{n-1}^{k-1}, \quad (7)$$

а с другой –

$$F_j = \delta_1 C_{n-1}^{k-q_1} + \tau_1. \quad (8)$$

Сравнивая (7), (8) и учитывая $F_j < C_n^k$, приходим к выводу, что $0 \leq \delta_1 \leq 1$, $q_1 = 0$ и $\tau_1 < C_{n-1}^{k-1}$.

Далее, производим деление остатка τ_1 на весовой коэффициент ρ_2 при $i=2$, т.е. $\tau_1 = \delta_2 \rho_2 + \tau_2$. При $i=2$ имеем $\rho_2 = C_{n-2}^{k-q_2}$. Преобразовывая (7) и (8) уже в виде

$$C_{n-1}^{k-1} = C_{n-2}^{k-1} + C_{n-2}^{k-2}, \quad \tau_1 = \delta_2 C_{n-2}^{k-q_2} + \tau_2,$$

делаем вывод, что $0 \leq \delta_2 \leq 1$, $q_2 = 1$ и $\tau_2 < C_{n-2}^{k-2}$.

На i -м шаге деления остатка τ_{i-1} выражения для суммы биномиальных коэффициентов и деления $(i-1)$ -го остатка будут иметь вид:

$$C_{n-i+1}^{k-i+1} = C_{n-i}^{k-i+1} + C_{n-i}^{k-i}, \quad \tau_{i-1} = \delta_i C_{n-i}^{k-q_i} + \tau_i,$$

на основании чего можно заключить, что при $\tau_{i-1} < C_{n-i+1}^{k-i+1}$ имеем $0 \leq \delta_i \leq 1$, $q_i = i-1$ и

$\tau_i < C_{n-i}^{k-i}$.

На $i = r - 1$ шаге имеем $\tau_{r-2} < C_{n-r+2}^{k-r+2}$, биномиальный коэффициент уже вида C_{n-r+2}^{k-r+2} разбиваем согласно свойству суммирования чисел сочетаний и осуществляем деление остатка τ_{r-2} на весовой коэффициент ρ_1 :

$$C_{n-r+2}^{k-r+2} = C_{n-r+1}^{k-r+2} + C_{n-r+1}^{k-r+1}, \tau_{r-2} = \delta_{r-1} C_{n-r+1}^{k-q_{r-1}} + \tau_{r-1}.$$

Исходя из приведенных выражений для шага $i = r - 1$, при $\tau_{r-2} < C_{n-r+2}^{k-r+2}$ получаем $0 \leq \delta_{r-1} \leq 1$, $q_{r-1} = r - 2$ и $\tau_{r-1} < C_{n-r+1}^{k-r+1}$.

На последнем шаге $i = r$ аналогично формируя выражения суммы биномиальных коэффициентов и деления остатка, получаем

$$C_{n-r+1}^{k-r+1} = C_{n-r}^{k-r+1} + C_{n-r}^{k-r}, \tau_{r-1} = \delta_r C_{n-r}^{k-q_r} + \tau_r.$$

На основании приведенных равенств можно заключить, что при $\tau_{r-1} < C_{n-r+1}^{k-r+1}$ имеем $0 \leq \delta_r \leq 1$, $q_r = r - 1$ и $\tau_r < C_{n-r}^{k-r}$.

Рассмотрим неравенство для последнего остатка $\tau_r < C_{n-r}^{k-r}$. Переменная k представляет собой максимальное количество включений весовых коэффициентов вида $\rho_i = C_{n-i}^{k-q_i}$ в число F_j . Очевидно, если r – количество этапов деления исходного числа F_j , то $r = k$. Отсюда, $\tau_r < C_{n-r}^{k-r} = C_{n-r}^{k-k} = C_{n-r}^0 = 1$, а значит $\tau_r = 0$.

Выполнив последовательную подстановку выражений для τ_i в равенства для τ_{i-1} , получим представление требуемого вида:

$$F_j = \delta_1 C_{n-1}^k + \delta_2 C_{n-2}^{k-q_2} + \dots + \delta_i C_{n-i}^{k-q_i} + \dots + \delta_{r-1} C_{n-r+1}^{k-q_{r-1}} + \delta_r C_{n-r}^{k-q_r} = \sum_{i=1}^r \delta_i C_{n-i}^{k-q_i},$$

где $0 \leq \delta_i \leq 1$, q_i – сумма δ_t от первого до $(i-1)$ -го шага включительно, т.е. $q_i = \sum_{t=1}^{i-1} \delta_t$.

Таким образом, в результате приравнивания $\delta_i = x_i$ существование разложения (6) доказано.

Докажем теперь единственность разложения (6). Пусть при неизменных параметрах n и k номеру F_j соответствуют два биномиальных числа $X'_j = (x'_1 x'_2 \dots x'_i \dots x'_r)$ и $X''_j = (x''_1 x''_2 \dots x''_i \dots x''_r)$, которые имеют разложения (6):

$$F_j = F' = \sum_{i=1}^{r'} x'_i C_{n-i}^{k-q'_i} \quad \text{и} \quad F_j = F'' = \sum_{i=1}^{r''} x''_i C_{n-i}^{k-q''_i}, \quad (9)$$

причем $0 \leq x'_i \leq 1$ и $0 \leq x''_i \leq 1$.

Биномиальные числа X'_j и X''_j можно разбить на общие $X_j'^g, X_j''^g$ и собственные $X_j'^s, X_j''^s$ части [5, 9]. При этом для соответствующих номеров F' и F'' , вычисленных согласно (9), должны выполняться равенства $F' = \tilde{X}_j'^g + \tilde{X}_j'^s$ и $F'' = \tilde{X}_j''^g + \tilde{X}_j''^s$, где знаком “ \sim ” обозначены количественные эквиваленты общих и собственных кодовых частей чисел X'_j и X''_j . Очевидно, для общих частей $\tilde{X}_j'^g = \tilde{X}_j''^g$. Тогда $F'' - F' = \tilde{X}_j''^s - \tilde{X}_j'^s$ и с учетом нашего предположения эта разность должна быть равной нулю.

Предположим, что собственные части $X_j''^s$ и $X_j'^s$ отличаются друг от друга максимальным образом, имеют противоположные значения соответствующих разрядов, начиная со старших. Пусть $X_j''^s = (100\dots 0)$ и $X_j'^s = (011\dots 1)$. Но согласно [5, 9] для такого случая $F'' - F' = \tilde{X}_j''^s - \tilde{X}_j'^s = 1$, т.е. $F'' \neq F'$. Из [5, 9] также следует, что $X_j'^s = (011\dots 1)$ достигает своего максимального значения. Поэтому появление одного или нескольких нулей вместо единиц в $X_j'^s = (011\dots 1)$ изменит величину $\tilde{X}_j'^s$ в меньшую сторону и соответственно разность $F'' - F'$ станет еще больше единицы.

Появление одной или нескольких единиц вместо нулей в $X_j^{ns} = (100\dots 0)$ приведет к увеличению \tilde{X}_j^{ns} и соответственно к увеличению разности $F'' - F'$, которая станет также больше единицы.

Одновременное нахождение нулей и единиц в старших разрядах в кодовых частях X_j'' и X_j' запрещено, поскольку они тогда будут относиться к общей части разложений F'' и F' . Других вариантов построения собственных частей X_j^{ns} и X_j^{ns} с тем, чтобы добиться равенства нулю разности $F'' - F'$, нет.

В результате при любых возможных значениях цифр в разрядах собственных частей X_j^{ns} и X_j^{ns} будет наблюдаться неравенство $\tilde{X}_j^{ns} \neq \tilde{X}_j^{ns}$ и, следовательно, $F'' \neq F'$. Тогда при неизменных параметрах n и k номеру F_j не могут соответствовать два разных биномиальных числа $X_j' = (x_1'x_2'\dots x_r')$ и $X_j'' = (x_1''x_2''\dots x_r'')$, т.е. $x_i' = x_i''$, $r' = r''$ и отсюда $X_j'' = X_j'$. Так как $X_j'' = X_j' = X_j$, то можно заключить, что разложение (6) является единственным для номера F_j . Теорема 3 доказана.

Если в теореме 3 вместо отображения $\phi^{-1} : Y[(k-1, k)] \rightarrow X[n, k]$ включить отображение вида $\phi^{-1} : Y[(k)] \rightarrow X[n, k]$ для кодов-сочетаний с ограничением $R_Y = (k)$, то получим теорему о перечислении равновесных кодов, причем суть доказательства второго этапа отображений останется прежней.

Теорема 4. О генерировании кодов-сочетаний с ограничением $R_Y = (k-1, k)$ (квази-равновесных кодов).

Обратное преобразование номера $F_j \in F = \{0, 1, 2, \dots\}$ в последовательность $Y_j \in Y[(k-1, k)]$, $Y_j = (y_1y_2\dots y_i\dots y_{n-1})$, $y_i \in \{0, 1\}$, удовлетворяющую ограничению $R_Y = (k-1, k)$, осуществляется с помощью отображений:

1) *отображения $\psi^{-1} : F \rightarrow X[n, k]$ номера F_j на биномиальное число $X_j \in X[n, k]$, $X_j = (x_1x_2\dots x_i\dots x_r)$, $x_i \in \{0, 1\}$ в соответствии с рекуррентным соотношением:*

$$x_i = \text{sign}(\text{sign}(F(i) - \rho_i) + 1), \quad (10)$$

где $F(1) = F_j$, $F(i) = F(i-1) - x_i\rho_i$, $\rho_i = C_{n-i}^{k-q_i}$;

2) *отображения $\phi : X[n, k] \rightarrow Y[(k-1, k)]$ биномиального числа X_j на искомую последовательность Y_j .*

В целом доказательство теоремы 4 проводится аналогичным способом, как это делается в теореме 3, учитывая, что обоснованием второго этапа генерирования кодов-сочетаний с ограничением $R_Y = (k-1, k)$ служит теорема 1.

5. Выводы

Научная новизна и практическая значимость. В данной работе заложены математические основы перечисления и генерирования кодов-сочетаний на основе биномиальных отображений, позволяющие строить математические модели указанных информационных процессов для ИУС. На базе разработанных моделей перечисления и генерирования возможна более эффективная реализация различных информационных задач ИУС, например, помехоустойчивой передачи и контроля ошибок аппаратуры, коммутации сообщений и каналов передачи, оптимизации процессов обработки данных. Используя то, что в структуре множеств кодов-сочетаний обнаружены биномиальные числа, также достаточно эффективно можно решать задачи вычислительной обработки данных ИУС, представленных рассматриваемыми комбинациями. В результате применения построенных моделей в ИУС затраты как временные, так и аппаратно-программные на выполнение информационных задач, использующих коды-сочетания, будут уменьшаться, что и соответствует сформулированной цели данной работы.

Перспективой будущих исследований в представленном направлении является дальнейшее развитие полученных математических моделей в целях охвата более широкого класса кодов-сочетаний, удовлетворяющих различного рода ограничениям, а также разработка прикладных алгоритмов для решения конкретных задач ИУС.

Список литературы: 1. *Амелькин В.А.* Методы нумерационного кодирования. Новосибирск: Наука, 1986. 155 с. 2. *Крюков А.Б.* Кодирование дискретных монотонных функций. В кн.: Аппаратура для космических исследований. М.: Наука, 1973. с. 10-15. 3. *Рейнгольд Э., Нивергельд Ю.* Комбинаторные алгоритмы. Теория и практика. М.: Мир, 1980. 476 с. 4. *Kreher, D., Stinson, D.* Combinatorial Algorithms: Generation, Enumeration and Search. CRC Press, 1999. 329 p. 5. *Борисенко А.А.* Биномиальный счет. Теория и практика: Монография. Сумы: ИТД “Университетская книга”, 2004. 170 с. 6. *Cover, T.* Enumerative Source Encoding / IEEE Transactions on Information Theory, Vol. IT-19, No. 1. Jan. 1973. pp. 73-77. 7. *Кувырков П.П., Темников Ф.Е.* Комбинаторные системы. М.: Энергия, 1975. 152 с. 8. *Кофман А.* Введение в прикладную комбинаторику. М.: Наука, 1975. 317 с. 9. *Борисенко А.А., Кулик И.А.* Биномиальное кодирование: Монография. Сумы: Изд-во СумГУ, 2010. 206 с. 10. *Kohzuki K., Tokiwa K., Tanaka H.* A Class of Single Error Correcting Constant Weight Codes / Electronics and Communications in Japan, Part 3. 1997. Vol. 80, No. 7. P. 55-64. 11. *Smith, D. et al.* Application of Coding Theory to the Design of Frequency Hopping Lists. Technical Report UG-M-02-1, University of Glamorgan, Wales, UK, 2002. 148 p. 12. *Кулик И.А., Скордина Е.М., Костель С.В.* Формирование квазиравновесных кодов на основе двоичных биномиальных чисел / Вісник СумДУ. 2010. № 1. С. 134-142. 13. *Кулик И.А., Скордина Е.М.* Биномиальные преобразования квазиравновесных кодов / Вісник СумДУ. 2010. № 3. С. 178-186.

Поступила в редколлегию 04.06.2011

Кулик Игорь Анатольевич, канд. техн. наук, докторант кафедры электроники и компьютерной техники Сумского государственного университета. Научные интересы: разработка специальных систем кодирования и на их основе устройств сжатия и передачи информации. Адрес: Украина, 40007, Сумы, ул. Римского-Корсакова, 2, тел. (0542) 33-55-39. e-mail: kulyk@ekt.sumdu.edu.ua.

Скордина Елена Михайловна, аспирантка кафедры электроники и компьютерной техники Сумского государственного университета. Научные интересы: адаптивные системы передачи данных, комбинаторное кодирование. Адрес: Украина, 40007, Сумы, ул. Римского-Корсакова, 2, тел. (0542) 33-55-39. e-mail: e.skordina@ekt.sumdu.edu.ua.

Костель Сергей Викторович, ассистент кафедры электроники и компьютерной техники Сумского государственного университета. Научные интересы: адаптивные системы сжатия информации, комбинаторное кодирование. Адрес: Украина, 40007, Сумы, ул. Римского-Корсакова, 2, тел. (0542) 33-55-39. e-mail: eikt@ukr.net.

УДК 681.518:004.93.1'

А.С. ДОВБИШ, О.Б. БАРИЛО

РОЗПАРАЛЕЛЕНИЙ АЛГОРИТМ ОПТИМІЗАЦІЇ ПАРАМЕТРІВ НАВЧАННЯ СППР ДЛЯ КЕРУВАННЯ СЛАБОФОРМАЛІЗОВАНИМ ТЕХНОЛОГІЧНИМ ПРОЦЕСОМ

Розглядається інформаційно-екстремальний алгоритм навчання системи підтримки прийняття рішень з мультипоточною паралельною оптимізацією системи контрольних допусків на ознаки розпізнавання, що мають різні шкали вимірювання. Це дозволяє підвищити оперативність навчання та визначити оптимальні значення параметрів за глобальним максимумом інформаційного критерію функціональної ефективності в робочій області визначення його функції.

Вступ

Застосування сучасних засобів автоматизації у хімічній промисловості виявляється недостатньо ефективним як через довільні початкові умови технологічних процесів, так і через часові затримки результатів хімічного аналізу, що ускладнює організацію неперервного вхідного, міжопераційного та вихідного контролю. Це обумовлює на виході відносно високий відсоток некондиційного продукту. Одним із способів підвищення функціональної ефективності автоматизованої системи керування (АСК) є надання їй властивості адаптив-

ності на основі машинного навчання та розпізнавання образів [1]. Для цього перспективним є використання ідей і методів інформаційно-екстремальної інтелектуальної технології (ІЕІ-технологія), що ґрунтується на максимізації інформаційної спроможності інтелектуальної АСК у процесі її навчання [1,2].

В [1] розглядалися у рамках ІЕІ-технології питання паралельної оптимізації контрольних допусків на ознаки, які мали однакову шкалу вимірювання. Але в загальному випадку в завданнях класифікаційного керування, наприклад, розподіленими технологічними процесами технологічні параметри мають різні шкали вимірювання, а сам процес оптимізації є досить громіздкий та потребує значну кількість ресурсів обчислювальної техніки та часу. В даній роботі запропоновано розпаралелений між вузлами обчислення [3,4] інформаційно-екстремальний алгоритм навчання СППР за паралельною оптимізацією контрольних допусків на ознаки розпізнавання з різними шкалами вимірювання.

1. Постановка задачі

Необхідно у рамках ІЕІ-технології реалізувати базовий алгоритм навчання та паралельну оптимізацію контрольних допусків на ознаки, які мають різні шкали вимірювання в рамках технологічного процесу. Навчальна матриця складається з 4 класів, кожен з яких має 89 ознак розпізнавання та 600 реалізацій. Оскільки навчальні матриці є досить об'ємні, а сам процес оптимізації досить громіздкий та потребує значну кількість ресурсів обчислювальної техніки і часу, то пропонується розпаралелити алгоритм оптимізації для проведення обчислень на багатоядерній обчислювальній техніці.

Нехай для алфавіту нечітких класів розпізнавання $\{X_m^0 \mid m = \overline{1, M}\}$, які характеризують M допустимих функціональних станів складного організаційно-технічного комплексу, сформовано апріорно класифіковану нечітку навчальну матрицю типу «об'єкт-властивість» $\|y_{m,i}^{(j)}\|$, $i = \overline{1, N}$, $j = \overline{1, n}$, де N, n – кількість ознак розпізнавання і випробувань відповідно. Крім того, дано вектор параметрів функціонування СППР $g = \langle x_m, d_m, \delta \rangle$, де x_m, d_m, δ – еталонний вектор-реалізація класу X_m^0 , радіус контейнера класу X_m^0 , що відновлюється в радіальному базисі простору ознак розпізнавання, і параметр поля контрольних допусків відповідно. При цьому задано такі обмеження: x_m – вектор, вершина якого визначає геометричний центр контейнера класу X_m^0 , $d_m \in [0; d(x_m \oplus x_c) - 1]$, де $d(x_m \oplus x_c)$ – кодова відстань центру класу X_m^0 від центру найближчого (сусіднього) до нього класу X_c , і параметр $\delta \in [0; \delta_H / 2]$, де δ_H – нормоване (експлуатаційне) поле допусків для відносної шкали вимірювання ознак, яке є областю значень для параметра контрольного поля допусків δ . При цьому нижній допуск на i -ту ознаку дорівнює $A_{H_1} = y_{1,i} - \delta$, а верхній допуск – $A_B = y_{1,i} + \delta$. Тут $y_{1,i}$ – i -та ознака еталонного вектора-реалізації y_1 базового класу X_1^0 , який характеризує найбільш бажаний для особи, що приймає рішення, функціональний стан СППР.

На етапі навчання необхідно оптимізувати параметри навчання за усередненим інформаційним критерієм функціональної ефективності (КФЕ):

$$E = \frac{1}{M} \sum_{m=1}^M E_m^* \quad (1)$$

де E_m^* – глобальний максимум інформаційного КФЕ навчання СППР розпізнавати реалізації класу X_m^0 і на етапі екзамену приймати високостовірне рішення про належність вектора-реалізації образу, що розпізнається, до деякого класу із заданого алфавіту $\{X_m^0\}$.

2. Опис алгоритму

Ідея алгоритму навчання у рамках ІЕІ-технології полягає в оптимізації просторово-часових параметрів функціонування СППР шляхом пошуку глобального максимуму інформаційного критерію (1) в робочій (допустимій) області визначення його функції. При цьому на кожному кроці навчання відбувається відновлення в дискретному просторі ознак опти-

мальних контейнерів класів розпізнавання за алгоритмом паралельної оптимізації контрольних допусків, який здійснюється за ітераційною процедурою

$$\delta^* = \arg \max_{G_\delta} \{ \max_{G_d \cap G_E} E \}, \quad (2)$$

де G_δ, G_d, G_E – допустимі області значень параметра поля допусків δ , радіусів контейнерів класів розпізнавання й інформаційного КФЕ (1) відповідно.

При реалізації ітераційного двочиклічного алгоритму (2) навчання СППР з паралельною оптимізацією контрольних допусків на ознаки розпізнавання внутрішній цикл реалізує базовий алгоритм навчання, який на кожному кроці обчислює значення критерію (1), здійснює в процесі навчання пошук глобального максимуму його функції і визначає оптимальні геометричні параметри контейнерів відповідних класів розпізнавання. Зовнішній цикл визначає оптимальне значення параметра δ , при якому КФЕ набуває максимального значення в робочій області визначення його функції. Вхідними даними є навчальна матриця $\|y_{m,i}^{(j)}\|$ та поточний параметр контрольного поля допусків δ . Розглянемо основні етапи алгоритму навчання з паралельною оптимізацією контрольних допусків на ознаки розпізнавання.

1. Формується еталонний вектор $y_1 \in X_1^0$ шляхом статистичного усереднення значень ознак в стовпчиках матриці $\|y_{1,i}^{(j)}\|$.

2. Обнуляється лічильник кроків зміни параметра δ .

3. $\delta := \delta + k$, де k – коефіцієнт збільшення кроку δ .

4. Для поточного значення параметра поля допусків δ обчислюються нижні та верхні контрольні допуски для всіх ознак: $\{A_{HK,i}[1] := y_{1,i} - \delta[1]\}$ і $\{A_{BK,i}[1] := y_{1,i} + \delta[1]\}$, $i = \overline{1, N}$ відповідно, де $y_{1,i}$ – вибіркове середнє значення i -ї ознаки для векторів-реалізацій класу X_1^0 , який є найбільш бажаним для особи, що приймає рішення.

5. Формується бінарна навчальна матриця $\|x_{m,i}^{(j)}\|$ за правилом

$$x_1^{(j)} = \begin{cases} 1, & \text{if } A_{HK,i} \leq y_{1,i} \leq A_{BK,i}; \\ 0, & \text{if else.} \end{cases} \quad (3)$$

6. Формується масив еталонних двійкових векторів-реалізацій $\{x_{m,i} \mid m = \overline{1, M}, i = \overline{1, N}\}$, елементи яких визначаються за правилом

$$x_{m,i} = \begin{cases} 1, & \text{if } \frac{1}{n} \sum_{j=1}^n x_{m,i}^{(j)} > \rho_m; \\ 0, & \text{if else,} \end{cases} \quad (4)$$

де ρ_m – рівень селекції координат вектора $x_m \in X_m^0$, за замовчуванням дорівнює $\rho_m = 0,5$.

7. Здійснюється розбиття множини еталонних векторів $\{x_m\}$ на пари найближчих «сусідів»: $\mathfrak{R}_m^{[2]} = \langle x_m, x_1 \rangle$, де x_1 – еталонний вектор сусіднього класу X_1^0 , за такою схемою:

а) структурується множина еталонних векторів, починаючи з вектора x_1 базового класу X_1^0 , який характеризує найбільшу функціональну ефективність;

б) будується матриця кодових відстаней між еталонними векторами розмірності $M \times M$;

в) для кожного рядка матриці кодових відстаней знаходиться мінімальний елемент, який належить стовпчику вектора, найближчого до вектора, що визначає рядок. За наявності декількох однакових мінімальних елементів вибирається з них будь-який, оскільки вони є рівноправними;

г) формується структурована множина елементів попарного розбиття $\{\mathfrak{R}_m^{[2]} \mid m = \overline{1, M}\}$.

8. Змінюється кодова відстань d_m і обчислюється інформаційний КФЕ.

9. Здійснюється пошук максимуму КФЕ в робочій області G_E його визначення: $E_m^* = \max_{\{d\}} E_m$, де $\{d\} = \{d_1, \dots, d_{\max}\} \in [0; d(x_m \oplus x_1) - 1]$ – множина радіусів концентрованих гіперсфер, центр яких визначається вершиною еталонного вектора $x_m \in X_m^0$. При цьому множина $\{d\}$ є так само множиною кроків навчання СППР.

10. Якщо $E_1^* [1] \in G_E$, то виконується пункт 9, інакше – пункт 8.

11. Якщо $\delta \leq \delta_H / 2$, то виконується пункт 3, інакше – пункт 12.

12. $\{A_{НК,i}^* := A_{НК,i}[1-1]\}; \{A_{ВК,i}^* := A_{ВК,i}[1-1]\}, i = \overline{1, N}; E_1^* := E_1^*[1-1]$.

13. ЗУПИНКА.

Як КФЕ навчання використовується одна із статистичних інформаційних мір. Наприклад, для двохальтернативних рішень модифікований критерій Кульбака, що використовується для оцінки функціональної ефективності навчання СППР розпізнавати реалізації класу X_m^0 , має вигляд [2]

$$E_m^{(k)} = \log_2 \left(\frac{2 - (\alpha_m^{(k)}(d) + \beta_m^{(k)}(d))}{\alpha_m^{(k)}(d) + \beta_m^{(k)}(d)} \right) * [1 - (\alpha_m^{(k)}(d) + \beta_m^{(k)}(d))], \quad (6)$$

де $\alpha_m^{(k)}(d), \beta_m^{(k)}(d)$ – точнісні характеристики: похибки першого та другого роду відповідно, що обчислюються на k -му кроці навчання; d – дистанційна міра, що визначає радіуси контейнерів, побудованих у радіальному базисі простору ознак розпізнавання.

Оцінками похибок першого та другого роду при прийнятті рішень в процесі навчання є емпіричні частоти

$$\tilde{\alpha}_m^{(k)}(d) = \frac{K_{1,m}^{(k)}}{n_{\min}}, \quad \tilde{\beta}_m^{(k)}(d) = \frac{K_{2,m}^{(k)}}{n_{\min}}, \quad (13)$$

де $K_{1,m}^{(k)}$ – кількість подій, що визначають неналежність реалізацій образу контейнера класу X_m^0 , якщо насправді вони є реалізаціями цього класу; $K_{2,m}^{(k)}$ – кількість подій, що визначають належність реалізацій образу контейнера класу X_m^0 , якщо вони належать іншому класу [1].

При використанні багатоядерної обчислювальної техніки пропонується розділити зовнішній цикл визначення оптимального значення параметра δ на рівні частини відповідно до кількості вузлів обчислення [4, 5] (у нашому випадку передбачається використання 2-ядерного процесора, тому зовнішній цикл для першого ядра буде мати діапазон $0 < \delta \leq \delta_H / 4$, для другого ядра: $\delta_H / 4 < \delta \leq \delta_H / 2$).

3. Приклад реалізації алгоритму розпаралеленого навчання

Розглянемо застосування наведеного вище алгоритму навчання СППР з розпаралеленою оптимізацією контрольних допусків на ознаки розпізнавання для керування технологічним процесом виробництва фосфорної кислоти у ВАТ «Сумихімпром» (м. Суми). Вхідну апріорно класифіковану навчальну матрицю було сформовано за експериментальними даними, одержаними для чотирьох класів, що характеризували зміст P_2O_5 на виході п'ятого ступеня екстрактора. При цьому кількість ознак розпізнавання дорівнювала 89, 15 з яких надходили за результатами хімічного аналізу, а інші значення різних за природою ознак розпізнавання надходили в СППР безпосередньо з датчиків інформації. На рис. 1 показано графіки залежності КФЕ (6) від радіусів гіперсферичних контейнерів класів розпізнавання, одержані при реалізації базового алгоритму навчання TEACHING при стартовій системі контрольних допусків $\delta = \pm 14\%$ від номінального (усередненого) значення відповідної ознаки.

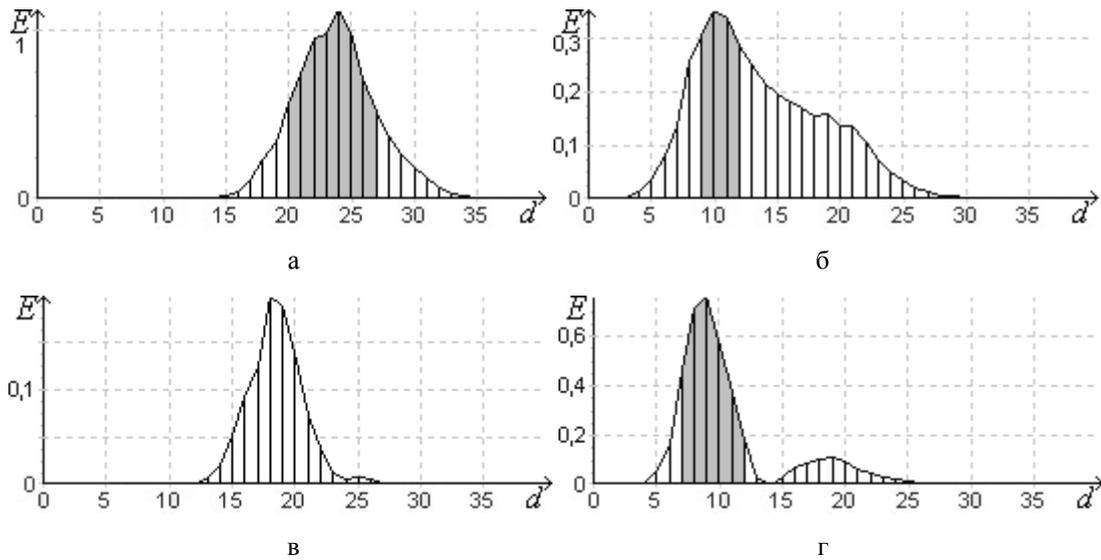


Рис. 1. Графіки залежності критерію Кульбака від радіусів гіперсферичних контейнерів класів розпізнавання: а – клас X_1^0 ; б – клас X_2^0 ; в – клас X_3^0 ; г – клас X_4^0

На рис. 1 і далі темні ділянки графіків означають робочі (допустимі) області, в яких значення достовірностей перебільшують відповідні помилки першого та другого роду. Аналіз рис. 1 показує, що усереднене максимальне значення КФЕ дорівнює $\bar{E}_{\max} = 0,56$. При цьому на графіку для третього класу відсутня робоча область визначення інформаційного критерію, що не дозволяє здійснювати оцінку функціональної ефективності навчання у рамках ІЕІ-технології. З метою усунення цього недоліку і збільшення усередненого КФЕ навчання СППР було реалізовано алгоритм паралельної оптимізації контрольних допусків на ознаки розпізнавання за схемою (2). На рис. 2 показано графік залежності критерію Кульбака від параметра поля контрольних допусків на ознаки розпізнавання δ .

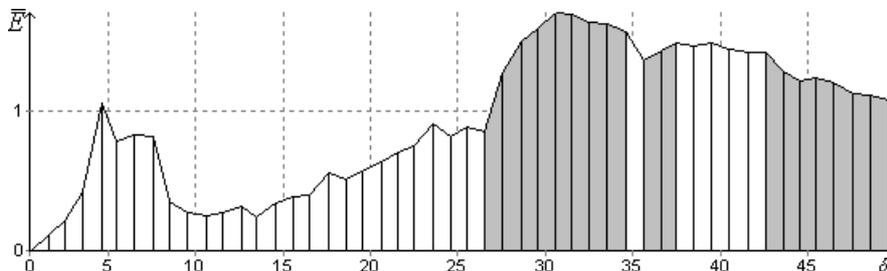


Рис. 2. Графіки залежності усередненого значення критерію Кульбака від величини поля контрольних допусків (у відсотках)

Аналіз рис. 2 показує, що графік має декілька робочих областей, в яких визначаються глобальні максимальні значення усередненого КФЕ. При цьому за максимальним глобальним значенням КФЕ оптимальне значення параметра поля контрольних допусків дорівнює $\delta = \pm 31\%$.

На рис. 3 показано графіки залежності КФЕ (б) від радіусів гіперсферичних контейнерів класів розпізнавання, одержані при оптимальному параметрі поля допусків на ознаки розпізнавання.

Аналіз рис. 3 показує, що після оптимізації контрольних допусків усереднене максимальне значення КФЕ збільшилося і дорівнює $\bar{E}_{\max} = 1,85$ при оптимальних значеннях радіусів контейнерів $d_1 = 13$ (тут і далі у кодових одиницях), $d_2 = 14$, $d_3 = 12$, $d_4 = 12$.

У результаті застосування розпаралеленого мультипоточкового алгоритму оптимізації кроку δ час роботи алгоритму зменшився до 4633мс порівняно з однопоточною оптимізацією даного параметра, що дорівнює 6724мс, збільшивши таким чином оперативність на 31,1%.

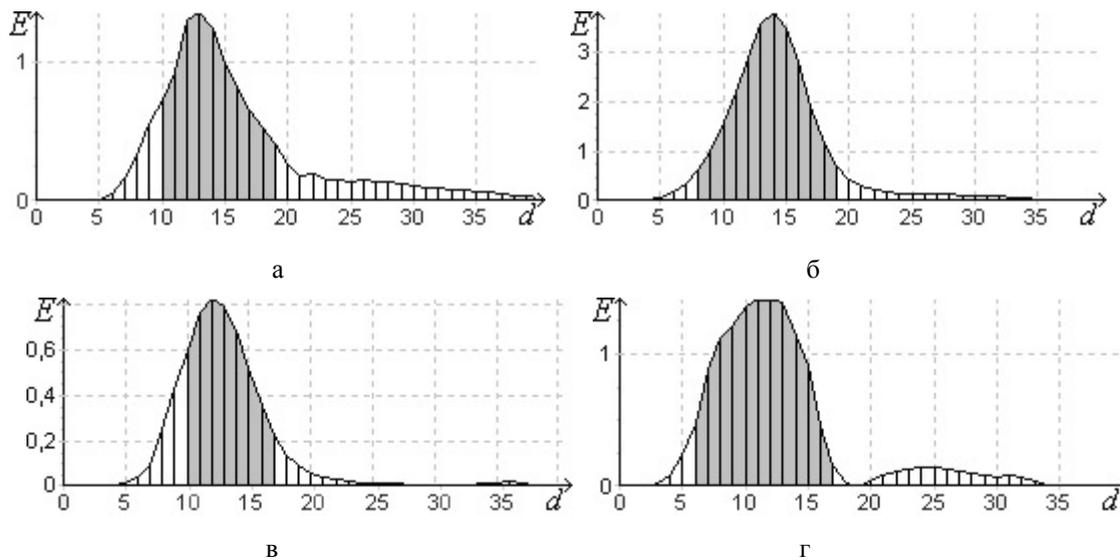


Рис. 3. Графіки залежності критерію Кульбака від радіусів контейнерів класів розпізнавання при оптимізації контрольних допусків на ознаки розпізнавання:
 а – клас X_1^0 ; б – клас X_2^0 ; в – клас X_3^0 ; г – клас X_4^0

Висновки

1. Запропонований інформаційно-екстремальний алгоритм паралельної оптимізації системи контрольних допусків на ознаки розпізнавання, що мають різні шкали вимірювання, дозволяє побудувати оптимальне нечітке розбиття простору ознак на класи розпізнавання.

2. Оптимізація системи контрольних допусків на ознаки розпізнавання у рамках ІЕІ-технології є необхідною умовою побудови безпомилкових за навчальною матрицею вирішальних правил, оскільки гарантує наявність робочої області визначення функції інформаційного критерію.

3. Використання алгоритму мультипоточної оптимізації параметрів функціонування в поєднанні з принципом відкладених рішень дозволяє підвищити оперативність навчання та визначити оптимальні значення параметрів за глобальним максимумом інформаційного критерію функціональної ефективності в робочій області визначення його функції.

Список літератури: 1. *Краснопоясовський А. С.* Інформаційний синтез інтелектуальних систем керування, що навчаються: Підхід, що ґрунтується на методи функціонально-статистичних випробувань. Суми: Вид. СумДУ, 2003. 257 с. 2. *Довбиш А. С.* Основи проектування інтелектуальних систем: Навчальний посібник. Суми: Вид-во СумДУ, 2009. 171 с. 3. *David A. Patterson and John L. Hennessy.* Computer Organization and Design (Second Edition) Morgan Kaufmann Publishers, 1998. ISBN 1558604286. P. 715. 4. *Asanovic et al:* Power is free, but transistors are expensive. New is power is expensive, but transistors are «free». 5. *Основи параллельного программирования / К.Ю. Богачев.* М.: БИНОМ. Лаборатория знаний. 2003. 342с. 6. *Основи многопоточного, параллельного и распределенного программирования.* Пер. с англ. М.: Издательский дом «ВИЛЬЯМС», 2003. 512с.

Надійшла до редколегії 30.05.2011

Довбиш Анатолій Степанович, д-р техн. наук, професор, завідувач кафедри інформатики Сумського державного університету. Наукові інтереси: інформаційний аналіз і синтез інтелектуальних систем, що навчаються. Адреса: Україна, 40007, Суми, ул. Римського-Корсакова, 2, тел. 050-307-83-74, e-mail: kras@id.sumdu.edu.ua.

Баріло Олександр Борисович, аспірант кафедри комп'ютерних наук Сумського державного університету. Наукові інтереси: інформаційний аналіз і синтез інтелектуальних систем, що навчаються (самонавчаються). Адреса: Україна, 40007, Суми, ул. Римського-Корсакова, 2, тел. 050-407-03-86, e-mail: alex.barylo@gmail.com.

РАЗРАБОТКА МЕТОДИК И УСТРОЙСТВ КОНТРОЛЯ СТРУКТУРНОГО СОВЕРШЕНСТВА КРЕМНИЕВЫХ СТРУКТУР

Предлагаются методики и устройства контроля структурного совершенства кремниевых эпитаксиальных структур. Показывается, что деформации структур, которые возникают в процессе шлифовки или вследствие значительных температурных градиентов, приводят к появлению внутренних напряжений и областей с высокой плотностью дислокаций. Предполагаемые авторами методы и устройства позволяют контролировать величину деформации КЭС и их зависимость от кристаллографических направлений. Делается вывод о том, что высокие температурные градиенты приводят к потере устойчивости КЭС и образованию у неё «антикластического» изгиба.

1. Введение

Проблеме исследования напряжений и деформаций в кремниевых эпитаксиальных структурах (КЭС) не уделялось достаточно большое внимание из-за отсутствия надежных высокопроизводительных методов контроля величины и характера сложных форм внутренних напряжений и деформаций. Деформации изгиба (выпучивание) появляются в пластинах, поверхность которых обрабатывается абразивными порошками, при образовании эпитаксиальных или диффузионных слоев, при наличии значительных температурных градиентов. Это приводит к появлению на поверхности КЭС значительных областей с недопустимо высоким уровнем плотности дислокаций, других структурных нарушений, приводящих к дальнейшему браку в производстве интегральных микросхем и других полупроводниковых приборов.

Из сказанного можно сделать вывод, что разработка моделей расчета внутренних напряжений и сложных форм деформаций, а также методик контроля этих параметров при производстве КЭС является весьма актуальной задачей.

2. Анализ проблемы и постановка задачи.

Известно, [1] что если пластина деформируется малыми усилиями, то до некоторого значения усилий её плоская форма является единственно возможной, а равновесие устойчивым. При критическом напряжении первоначальная плоская форма может оказаться неустойчивой и под действием ничтожно малых возмущений пластина перейдет к новой устойчивой форме равновесия с искривлённой срединной поверхностью.

Если принять, что W - изгиб пластины толщиной h , то радиусы изгиба R_x и R_y могут быть определены из следующих соотношений [1]:

$$\frac{1}{R_x} = -\frac{\partial^2 W}{\partial x^2} \quad \text{и} \quad \frac{1}{R_y} = -\frac{\partial^2 W}{\partial y^2}. \quad (1)$$

С учетом [1] получим для плоскости (111) значение плоских напряжений:

$$\sigma_x = -\frac{E_{(111)}Z}{1-\nu_{(111)}^2} \left(\frac{1}{R_x} + \nu \frac{1}{R_y} \right), \quad (2)$$

$$\sigma_y = -\frac{E_{(111)}Z}{1-\nu_{(111)}^2} \left(\frac{1}{R_y} + \nu \frac{1}{R_x} \right)$$

и для плоскости (100) в направлении $\langle 100 \rangle$

$$\sigma_x = \frac{E_{\langle 100 \rangle (001)}Z}{1-\nu_{\langle 100 \rangle (001)}^2} \left(\frac{1}{R_x} + \nu \frac{1}{R_y} \right); \quad (3)$$

$$\sigma_y = \frac{E_{\langle 100 \rangle \langle 001 \rangle} Z}{1 - \nu^2_{\langle 100 \rangle \langle 001 \rangle}} \left(\frac{1}{R_y} + \nu \frac{1}{R_x} \right).$$

Введем жесткость пластины из [1], которая определяется как

$$D = -\frac{Eh^3}{12(1-\nu^2)}. \quad (4)$$

Введем также изгибающие моменты M_x и M_y :

$$M_x = -D \left(\frac{1}{R_x} + \nu \frac{1}{R_y} \right) = -D \left(\frac{\partial^2 W}{\partial x^2} + \nu \frac{\partial^2 W}{\partial y^2} \right), \quad (5)$$

$$M_y = -D \left(\frac{1}{R_y} + \nu \frac{1}{R_x} \right) = -D \left(\frac{\partial^2 W}{\partial y^2} + \nu \frac{\partial^2 W}{\partial x^2} \right).$$

Из условия равновесия $M_y=0$ получим из (5)

$$\frac{\partial^2 W}{\partial y^2} = -\nu \frac{\partial^2 W}{\partial x^2}, \quad (6)$$

т.е. пластина имеет в изогнутом состоянии две кривизны в противоположном направлении (антикластический изгиб).

В начальном, безмоментном состоянии, внутренние тангенциальные силы уравновешиваются внешними усилиями, приложенными к пластине.

Для изотропной круглой термоупругой пластины (КЭС) дифференциальное уравнение, определяющее изгиб пластины, имеет вид [2]:

$$\nabla \nabla W + \mu^2 \nabla W = 0, \quad (7)$$

где

$$\mu = -\frac{\rho}{D}, \quad (8)$$

$$\rho = -\frac{F}{\pi d}, \quad \sigma_R = -\frac{\rho}{h}, \quad (9)$$

здесь F – сила, равномерно приложенная к периметру пластины; σ_R – радиальное (сжимающее) напряжение за счет внешней силы; d – диаметр пластины; ∇ – оператор Лапласа.

В результате решения уравнения (7) с введением граничных условий получим выражение для критического напряжения:

$$\sigma_n^{кр} = \frac{D}{d^2 h} \mu_{0n}^2. \quad (10)$$

Величина $\mu_{0n} = \mu_n d$ определяется из трансцендентного уравнения:

$$\mu_0 J_n(\mu_0) - (1-\nu) J_{n+1}(\mu_0) = 0. \quad (11)$$

Для пластины жестко закрепленной по контуру:

$$J_{n+1}(\mu_0) = 0. \quad (12)$$

Здесь J_n и J_{n+1} функции Бесселя действительного аргумента. Определяя из (11) и (12) величину μ_0 для коэффициента Пуассона $\nu = 0,358$ (плоскость кремния (111) из [3]) находим из графического решения уравнения (11) $\mu_0 = 2,077$, а из (12) $\mu_0 = 3,832$) и подставляя вместо D его значение (4), получаем:

$$\sigma_{кр} = \frac{E}{12(1-\nu^2)} \mu_0^2 \left(\frac{h}{d} \right)^2. \quad (13)$$

Критическое значение внутреннего напряжения может создаваться при наличии радиального градиента температур ΔT . Предположим, что при резком охлаждении вблизи края пластины температура понижается. В этом случае периферийный слой, температура которого отличается от температуры пластины на ΔT , выступает как внешняя сжимающая сила. Учитывая, что $\varepsilon = \alpha \Delta T$ и $\sigma = E\varepsilon$, и подставляя эти значения в (13), получаем:

$$\Delta T_{кр} = \frac{\sigma_{кр}}{\lambda E} = \frac{\mu_0^2}{12\lambda(1-\nu^2)} \left(\frac{h}{d}\right)^2. \quad (14)$$

Величины μ_0 были вычислены нами для плоскости (111), где $\mu_0 = 2,077$, а также для плоскости (100). При этом так как для анизотропной КЭС, изготовленной из кремниевой пластины (100), решение уравнения (7) отсутствует, мы рассмотрели два случая:

- 1) вся КЭС (100) имеет плоскость, совпадающую с плоскостью изотропной пластины, имеющей упругие константы $E_{\langle 100 \rangle}$ и $\nu_{\langle 100 \rangle}$;
- 2) вся КЭС (100) имеет жесткость, совпадающую с жесткостью изотропной пластины, имеющей упругие константы $E_{\langle 110 \rangle}$ и $\nu_{\langle 110 \rangle}$.

Результаты расчетов приведены в таблице для $h = 0,002d$ и $\alpha_{1000^\circ\text{C}} = 1 \times 10^{-6}$.

Расчетные значения радиального градиента температур

КЭС	μ_0	$\mu_0^2 / (1 - \nu^2)$	$\Delta T_{кр} \text{ } ^\circ\text{C}$
(111)	2,077	4,94	41
(100), $E_{\langle 100 \rangle}$, $\nu_{\langle 100 \rangle}$;	2,034	4,58	37,3
(100), $E_{\langle 110 \rangle}$, $\nu_{\langle 110 \rangle}$;	1,9	3,63	31,6

Таким образом, видно, что для КЭС диаметром 60 мм при толщине $h = 300$ мкм. критическое охлаждение периферии, при котором КЭС теряет устойчивость, составляет при $1000^\circ\text{C} - 30 \div 40^\circ\text{C}$. До потери устойчивости градиент температуры уравнивается внутренними напряжениями сжатия, которые однородны по всей толщине пластины. Эти напряжения могут быть ниже предела пластичности и при дальнейшем остывании КЭС вплоть до комнатной температуры они исчезают, не оставляя никаких следов. При этом, в конечном счете, изгиб будет отсутствовать.

В момент потери устойчивости КЭС изгибается и вместо однородных по толщине напряжений сжатия возникают неоднородные по толщине напряжения изгиба, максимальная величина которых может превышать значения напряжения сжатия и, следовательно, может превышать предел пластичности. Возникает пластическая деформация, приводящая к сложным формам изгиба после окончания охлаждения КЭС.

Учитывая зависимость модуля Юнга E и коэффициента Пуассона ν от кристаллографических направлений в пластине кремния, ориентированной в плоскости (100), которая приведена на рис. 1 и построена нами по данным работы [3], а также зависимость $\sigma_{кр}$ и $T_{кр}$ от анизотропии упругих свойств и диаметра пластины, необходимо разработать методику прецизионного контроля изгибов КЭС в зависимости от кристаллографических направлений в плоскости (100), а также методику контроля внутренних напряжений в зависимости от кристаллографических направлений.

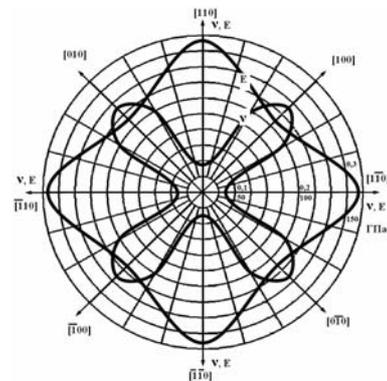


Рис. 1. Зависимость модуля Юнга E и коэффициента Пуассона ν от кристаллографических направлений в пластине кремния (100)

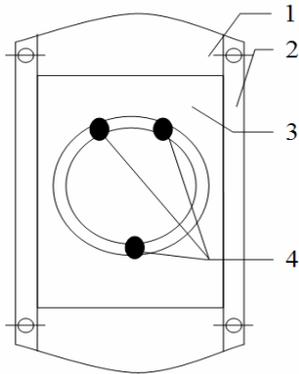


Рис. 1. Манипулятор для снятия топограмм: 1 – основание; 2 – направляющая; 3 – измерительный столик; 4 – корундовые наконечники

3. Разработка методики измерений и обсуждение результатов

Для измерения сложных форм изгиба КЭС и их зависимости от кристаллографических направлений в плоскости кремниевой пластины нами был разработан манипулятор, общий вид которого показан на рис.2. Основание манипулятора 1 вкручивается вместо штатного столика микроинтерферометра МИИ-4. Перемещение головки МИИ-4 контролируется с помощью индуктивно датчика перемещения, связанного с измерительным блоком (рис.3). Контролируемая КЭС устанавливается на измерительный столик 3 на корундовые наконечники 4, расположенные вдоль кристаллографического направления КЭС. Перемещая вручную КЭС вокруг корундовых наконечников на 360° , а с помощью столика 3 линейно по диаметру пластины, можно снять топограмму изгиба КЭС. Эпюры деформации КЭС в 3D формате приведены на рис.4. Ориентация КЭС – (100). С помощью разработанной нами [4] установки «Полярон-2», имеющей возможность по программе измерять сложные формы внутренних напряжений, нами были измерены внутренние напряжения в той же КЭС ориентации (100). Эпюры напряжений представлены на рис.5.

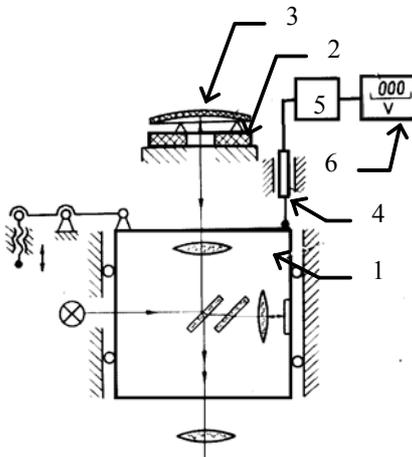


Рис. 2. Установка для снятия эпюр изгибов КЭС. 1 – микроинтерферометр МИИ4; 2 – измерительное кольцо; 3 – образец; 4 – индуктивный датчик; 5 – измерительный блок; 6 – индикаторное устройство

Как видно из рисунков, эксперимент подтвердил наши выводы о потере устойчивости КЭС и образовании у неё «антикластического» изгиба.

4. Выводы

1. На основании теории статической устойчивости тонких пластинок (что представляет собой КЭС) был предложен новый механизм, приводящий к новым формам изгиба КЭС ориентации (100).

2. Впервые были предложены выражения для определения критических напряжений $\sigma_{кр}$ и критического радиального градиента температур $T_{кр}$ и определена взаимосвязь этих выражений с диаметром и толщиной КЭС и упругими постоянными E и ν .

3. Экспериментальные исследования изгиба КЭС с ориентацией (100) показали, что в пластине (100) возникает сложный изгиб, эпюры которого имеют экстремум в одном из кристаллографических направлений $\langle 100 \rangle$,

тогда как КЭС, изготовленные из пластин ориентации $\langle 111 \rangle$, в большинстве случаев имеют чисто сферический изгиб.

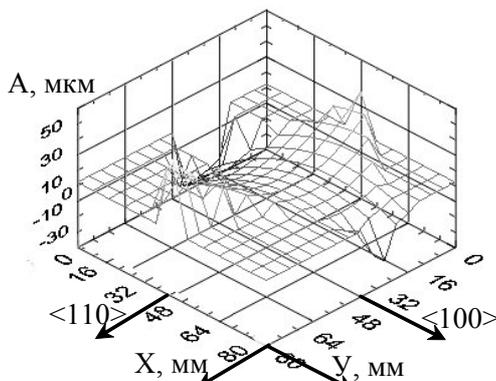


Рис. 4. Эпюры деформации пластины КЭС, вызванные «антикластическим» изгибом

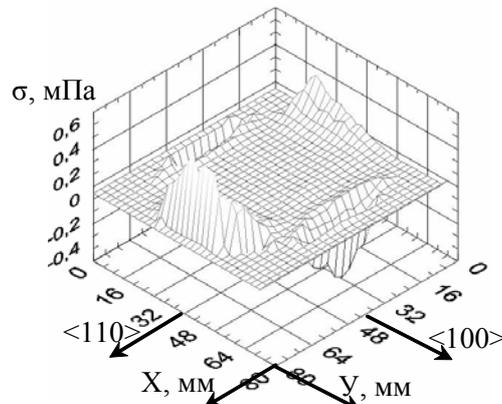


Рис. 5. Эпюры внутренних напряжений в КЭС

Список литературы: 1. Тимошенко С.П., Гудьер Дж. Теория упругости: Пер. с англ. под ред. Шапиро Г.С. М.: Наука, 1979. С. 34-53. 2. Амбарцумян С.А. Теория анизотропных пластин. М.: Наука, 1967. 227 с. 3. Wortman J.J. and Evans R.A. Young's modulus, shear's modulus and poisson's ratio in silicon and germanium // J. Appl. Phys. 1965. Vol. 36, №1. P. 153. 4. Оксанич А. П. Автоматизований комплекс вимірювання внутрішньої напруги в пластинах GaAs / А. П. Оксанич, С. Е. Притчин, В. Є. Краскевич, В. В. Батареев // Складні системи і процеси. 2006. №2 (10). С. 40-50.

Поступила в редколлегию 12.06.2011

Оксанич Анатолий Петрович, д-р техн. наук, профессор, зав. каф. информационно-управляющих систем, директор НИИ технологии полупроводников и информационно-управляющих систем Кременчугского национального университета им. Михаила Остроградского. Научные интересы: методы и аппаратура контроля структурно-совершенных полупроводниковых монокристаллов. Адрес: Украина, 39600, Кременчуг, ул. Первомайская, 20, тел.: (05366) 30157. E-mail: oksanich@kdu.edu.ua.

Притчин Сергей Эмильевич, канд. техн. наук, доцент кафедры информационно-управляющих систем Кременчугского национального университета им. Михаила Остроградского. Научные интересы: автоматизация процессов управления производством полупроводниковых материалов. Адрес: Украина, 39600, Кременчуг, ул. Первомайская, 20. E-mail: pritchin@knu.edu.ua.

Седин Евгений Александрович, старший преподаватель кафедры технической кибернетики Криворожского института экономики, информационных технологий и управления. Научные интересы: методы и аппаратура контроля структурно-совершенных полупроводниковых монокристаллов. Адрес: Украина, 50048, Кривой Рог, ул. Революционная, 5, тел. (097) 9002078. E-mail: e.a.sedin@gmail.com.

УДК 681.5+548.55

*М.В. ЗАГИРНЯК, А.П. ОКСАНИЧ, В.Р. ПЕТРЕНКО, С.Э. ПРИТЧИН,
В.А. ТЕРБАН*

РАЗРАБОТКА МАТЕМАТИЧЕСКОЙ МОДЕЛИ И ПРОГНОЗНОГО РЕГУЛЯТОРА ДЛЯ ПРОЦЕССА ВЫРАЩИВАНИЯ МОНОКРИСТАЛЛОВ ПОЛУИЗОЛИРУЮЩЕГО АРСЕНИДА ГАЛЛИЯ

На основе использования подхода Бокса-Дженкинса к синтезу моделей стохастических линейных динамических процессов разработана ARMAX-модель процесса вытягивания монокристаллических слитков полуизолирующего арсенида галлия по технологии с жидкостной герметизацией, связывающая вариации температуры фонового нагревателя с вариациями диаметра слитка. Полученная модель применяется для синтеза оптимального прогнозного регулятора процесса выращивания на этапе вытягивания цилиндрической части слитка. Результаты моделирования работы регулятора подтверждают его работоспособность.

1. Введение

Арсенид галлия привлекает внимание исследователей, как перспективный материал для микроэлектроники, благодаря ряду своих свойств. Важнейшим из них является высокая (в шесть раз большая, чем в кремнии) подвижность электронов в электрических полях низкой напряженности, что потенциально позволяет создать СВЧ - приборы с улучшенными характеристиками. Другое достоинство арсенида галлия заключается в большой ширине его запрещенной зоны, что является необходимым условием работоспособности структур при повышенных температурах. Кроме того, малая величина времени жизни неосновных носителей и большая, чем у кремния, ширина запрещенной зоны делают арсенид галлия более перспективным материалом для создания радиационно-стойких приборов и интегральных схем (ИС).

Однако выращивание крупногабаритных монокристаллов арсенида галлия, предназначенных для промышленного использования, сопряжено с большими материальными затра-

тами. В связи с этим особенно актуальной становится проблема повышения качества монокристаллов путем совершенствования систем управления.

Одним из основных параметров в процессе роста монокристаллов арсенида галлия является его диаметр. Обеспечение постоянства диаметра растущего кристалла по всей длине его цилиндрической части – самая актуальная задача управления процессом выращивания. Колебания диаметра во время выращивания приводят к неоднородности кристаллической решетки, что сказывается на качестве будущего полупроводникового прибора.

Г.А. Сатункин в своих работах [1,2] развил подход к управлению процессом кристаллизации на основе разработанной им математической модели низкого порядка (МНП), имеющей следующий вид в стандартной матричной форме:

$$X(t) = A(t)X(t) + B(t)\Omega(t), \quad (1)$$

где $X(t) = [\delta r(t), \delta h(t)]^T$ – вектор переменных состояния, компонентами которого являются малые отклонения радиуса кристалла и высоты мениска расплава от заданных значений; $A(t)$ и $B(t)$ – матричные коэффициенты, зависящие от температурных параметров объекта управления; $\Omega(t) = [\delta T(t), \delta V(t)]^T$ – вектор входных переменных, содержащий вариации температуры расплава и скорости вытягивания. Заметим, что МНП не учитывает в явном виде стохастичность процесса выращивания, хотя, как отмечено в [3], определенные перспективы в управлении процессом кристаллизации можно связывать именно с учетом стохастической природы технологического процесса получения монокристаллов.

В работе [4] исследовалась возможность использования для моделирования зависимости вариаций диаметра кристалла от вариаций скорости вытягивания класса ARMAX-моделей. Были получены положительные результаты. Следует также отметить, что использование ARMAX-моделей хорошо себя зарекомендовало в ряде работ по моделированию и управлению процессом выращивания монокристаллических слитков кремния [5-7].

Но промышленное использование регулятора, разработанного на основе модели [4] и описанного в [8], дало основание усомниться в оптимальности выбора канала управления. Оказалось, что малыми отклонениями скорости вытягивания слитка не удастся обеспечить требуемую точность регулирования диаметра, а при больших отклонениях наблюдается ухудшение качественных параметров кристаллов. Кроме того, существенные колебания скорости вытягивания могут приводить к срыву монокристаллическости роста слитка. По мнению авторов, избежать этого возможно, если для регулирования диаметра использовать канал фонового нагревателя.

Разработка ARMAX-модели, связывающей отклонения диаметра слитка с отклонениями температуры фонового нагревателя, и синтез на ее основе регулятора диаметра, соответствующего концепции прогнозного управления, и является *целью данной работы*. Метод прогнозного управления был разработан в конце семидесятых годов прошлого века [9]. На сегодня известно значительное количество модификаций метода прогнозного управления: EPSAC, GPC, MUSMAR, MAC, PFC, QDMC, SOLO [10-16]. В этих алгоритмах используется одна и та же концепция прогнозного управления: наличие внутренней модели, метод отступающего горизонта и вычисление последовательности прогнозных оптимальных сигналов управления. Алгоритмы отличаются применением различных моделей реальной системы, возмущающих воздействий и критериев оптимальности управления.

Все экспериментальные работы были выполнены на модернизированной промышленной технологической установке “Арсенид-1М” в условиях выращивания слитков арсенида галлия диаметром 40 мм.

2. Постановка задачи

Для построения математической модели будем использовать класс стохастических разностных моделей с дискретным временем [17]. Предположим, что модель передаточной функции

$$Y_t = v(B)X_t + N_t \quad (2)$$

может быть экономично параметризована в виде

$$Y_t = \delta^{-1}(B)\omega(B)X_{t-b} + N_t, \quad (3)$$

где b – чистое запаздывание; B – оператор сдвига назад, т.е. $BY_t = Y_{t-1}$;

$$\delta(B) = 1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_r B^r; \omega(B) = \omega_0 - \omega_1 B - \omega_2 B^2 - \dots - \omega_s B^s;$$

$Y_i, i = t, t-1, \dots, t-r$ и $X_j, j = t-b, t-b-1, \dots, t-b-s$ – соответственно отклонения выхода и входа от равновесных состояний; N_t - шум, генерируемый некоторым процессом авторегрессии-проинтегрированного скользящего среднего (АРПСС) и статистически независимый от X_t .

Таким образом, для построения модели необходимо по доступным наблюдениям $(X_1, Y_1), (X_2, Y_2), \dots, (X_N, Y_N)$ определить оценки параметров r, s, b и начальные оценки параметров $\delta_i, i = \overline{1, r}$ и $\omega_j, j = \overline{0, s}$, а также идентифицировать и оценить модель шума.

Для регулирования диаметра слитка будем использовать алгоритм прогнозного управления, который в общем виде может быть представлен последовательностью следующих шагов:

1. Предсказание в момент времени t (на основе модели объекта управления) значений выходной переменной системы $\hat{y}_t(t+k)$, где $k=1, \dots, N_1$. (При этом выход зависит от будущих управляющих воздействий $\hat{x}_t(t+k), k=0, 1, \dots, N_2$).

2. Выбор целевой функции управления и оптимизация с ее помощью $\hat{x}_t(t+k), k=0, 1, \dots, N_2$.

3. Реализация управления $x_t = \hat{x}_t(t)$.

4. Переход в момент времени $(t+1)$ к шагу 1 и повторение шагов 1–4 до достижения цели управления.

В качестве критерия оптимизации, используемого на шаге 2, возьмем критерий вида:

$$J(t) = \sum_{k=1}^{N_1} [\hat{y}_t(t+k) - r_t(t+k)]^2 q_1(k) + \sum_{k=0}^{N_2-1} [\hat{x}_t(t+k)]^2 q_2(k).$$

Значения штрафов на управление и его ошибку можно изменять с помощью весовых коэффициентов $q_2(k)$ и $q_1(k)$. Предполагается, что всегда $N_2 \leq N_1$ и что $\Delta \hat{x}_t(t+k) = 0$ для $k \geq N_2$.

3. Структурно-параметрический синтез математической модели

Заметим, что путем взятия конечных разностей над процессами X_t, Y_t модель (2) можно привести к виду

$$y_t = v_0 x_t + v_1 x_{t-1} + v_2 x_{t-2} + \dots + n_t, \quad (4)$$

где $y_t = \nabla^d Y_t, x_t = \nabla^d X_t, n_t = \nabla^d N_t$ - стационарные процессы с нулевыми средними значениями, d -порядок разности.

Процедура построения модели передаточной функции в соответствии с [17] сводится к выполнению следующих основных этапов:

– получение грубых оценок \hat{v}_j импульсного отклика в (4) с помощью алгоритма, основанного на выравнивании спектра входа;

– определение оценок $\hat{r}, \hat{s}, \hat{b}$ параметров r, s, b на основе анализа поведения последовательности \hat{v}_j ;

– вычисление начальных оценок $\hat{\delta}_i, i = \overline{1, r}$ и $\hat{\omega}_j, j = \overline{0, s}$ на основании оценок $\hat{v}, \hat{r}, \hat{s}, \hat{b}$;

– определение структуры и начальных оценок параметров модели шума;

– уточнение оценок параметров комбинированной модели;

– диагностическая проверка разработанной модели.

Для определения оценок \hat{v}_j параметров v_j модели (4) использовалось соотношение

$$\hat{v}_j = \frac{r_{\alpha\beta}(j) s_\beta}{s_\alpha}, \quad j=0, 1, 2, \dots, \quad (5)$$

где $r_{\alpha\beta}(j)$ – выборочная взаимная корреляционная функция процессов α_t и β_t , а s_β и s_α – выборочные стандартные отклонения для этих процессов.

Процесс α_t определяется путем подгонки АРСС-модели к процессу x_t , т.е.

$$\varphi_x(B)\theta_x^{-1}(B)x_t = \alpha_t, \quad (6)$$

а процесс β_t – результат применения преобразования $\varphi_x(B)\theta_x^{-1}(B)$ к процессу y_t , т.е.

$$\beta_t = \varphi_x(B)\theta_x^{-1}(B)y_t. \quad (7)$$

Модель (4) при этом может быть представлена в виде

$$\beta_t = v(B)\alpha_t + \xi_t, \quad (8)$$

где $\xi_t = \varphi_x(B)\theta_x^{-1}(B)\eta_t$.

При известных \hat{v}_j значения параметров r, s, b модели (3) можно оценить, используя следующие факты [17]: для модели вида (3) веса v_j импульсного отклика состоят из b нулевых значений v_0, v_1, \dots, v_{b-1} , последующих $s-r+1$ значений $v_b, v_{b+1}, \dots, v_{b+s-r}$ с произвольным поведением (таких значений нет, если $s < r$) и значений v_j при $j \geq b+s-r+1$, поведение которых определяется разностным уравнением r -го порядка с r начальными значениями $v_{b+s}, v_{b+s-1}, \dots, v_{b+s-r+1}$.

Начальные оценки параметров $\delta_i, i = \overline{1, r}$ и $\omega_j, j = \overline{0, s}$ определяются путем использования следующей системы уравнений:

$$\begin{aligned} v_j &= 0, \quad j < b; \\ v_j &= \delta_1 v_{j-1} + \delta_2 v_{j-2} + \dots + \delta_r v_{j-r} + \omega_0, \quad j = b; \\ v_j &= \delta_1 v_{j-1} + \delta_2 v_{j-2} + \dots + \delta_r v_{j-r} + \omega_{j-b}, \quad j = b+1, b+2, \dots, b+s; \\ v_j &= \delta_1 v_{j-1} + \delta_2 v_{j-2} + \dots + \delta_r v_{j-r}, \quad j > b+s. \end{aligned} \quad (9)$$

При известных оценках параметров передаточной функции можно найти оценки η_t с помощью соотношения

$$\hat{\eta}_t = y_t - \hat{\delta}^{-1}(B)\hat{\omega}(B)x_{t-b}. \quad (10)$$

Далее с помощью известных методов идентификации АРСС-процессов определяется структура модели для η_t и начальные оценки ее параметров.

На этапе оценивания модели решается задача одновременного эффективного оценивания параметров $b, \bar{\delta}, \bar{\omega}, \bar{\varphi}, \bar{\theta}$ ранее идентифицированной модели вида

$$y_t = \delta^{-1}(B)\omega(B)x_{t-b} + \varphi^{-1}(B)\theta(B)a_t. \quad (11)$$

Эта задача решается путем минимизации условной суммы квадратов

$$S_0(b, \bar{\delta}, \bar{\omega}, \bar{\varphi}, \bar{\theta}) = \sum_{t=u+p+1}^n a_t^2(b, \bar{\delta}, \bar{\omega}, \bar{\varphi}, \bar{\theta} | x_0, y_0, a_0), \quad (12)$$

где u – большее из r и $s+b$. Для поиска оценок параметров, минимизирующих (12), использовался известный метод Марквардта [18]. При этом ковариационная матрица оценок V определяется формулой

$$V = (X'X)^{-1}\hat{\sigma}_a^2, \quad (13)$$

здесь X – регрессионная матрица в линеаризованной модели, вычисленная на последней итерации в процедуре Марквардта, а $\hat{\sigma}_a^2$ – оценка остаточной дисперсии.

Диагностическая проверка соответствия комбинированной модели анализируемым данным основана на исследовании поведения остаточных ошибок \hat{a}_t , которые можно определить с помощью следующего соотношения:

$$\hat{a}_t = \hat{\theta}^{-1}(B)\hat{\varphi}(B)\left(y_t - \hat{\delta}^{-1}(B)\hat{\omega}(B)x_{t-b}\right), \quad (14)$$

где $t \geq u+1$, $u = \max\{r, s+b\}$ и все $\hat{a}_j, j = \overline{1, u+p}$ равны нулю.

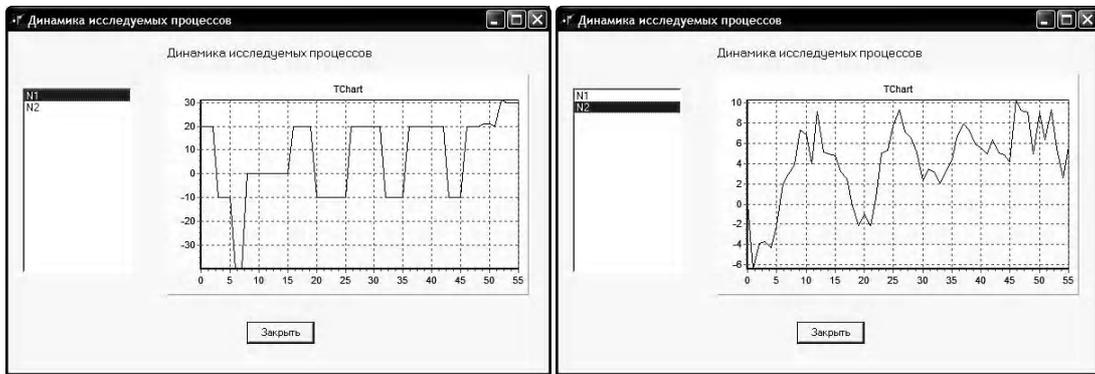
При этом вычисляется статистика

$$Q = m \sum_{k=1}^K r_{\hat{a}\hat{a}}^2(k), \quad (15)$$

где m – количество используемых в расчетах значений \hat{a}_t (обычно $m = n - u - p$, n – объем выборки); K – задержка, для которой справедливо, что при $k > K$ автокорреляции пренебрежительно малы.

В [17] указано, что величина Q распределена примерно как χ^2 с $K - p - q$ степенями свободы. Если Q меньше табличного значения для заданного уровня значимости, то принимается гипотеза об адекватности разработанной комбинированной модели.

Ниже приведены результаты основных этапов синтеза модели рассматриваемого класса, связывающей отклонения диаметра растущего кристалла с отклонениями температуры фонового нагревателя. В качестве исходных данных для решения задачи синтеза модели использовались два временных ряда, образованных наблюдаемыми значениями отклонений диаметра кристалла (Y_t) и отклонений температуры фонового нагревателя (X_t). Данные снимались на этапе выращивания цилиндрической части слитка с интервалом 1 мин на установке „Арсенид-1М”. Фрагменты рядов приведены на рис. 1.



а

б

Рис. 1. Фрагменты исследуемых рядов: а – отклонения температуры фонового нагревателя; б – отклонения диаметра слитка

Выборочная автокорреляционная и частная автокорреляционная функции входа x_t представлены на рис. 2 и 3 соответственно.

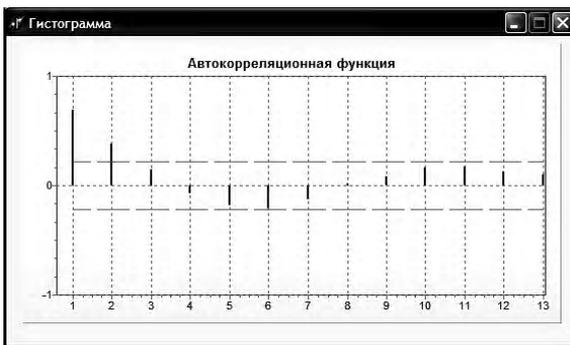


Рис. 2. Автокорреляционная функция

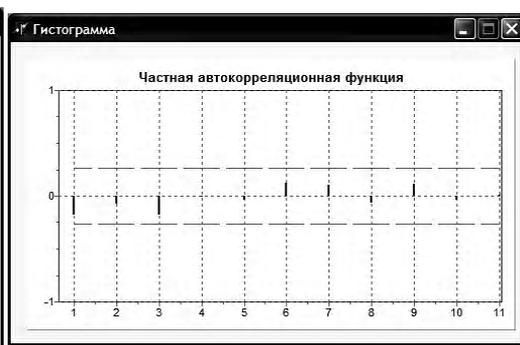


Рис. 3. Частная автокорреляционная функция

Анализ автокорреляционной и частной автокорреляционной функций позволяет предположить, что вход системы может быть представлен моделью скользящего среднего второго порядка. Для оценивания параметров модели входа x_t использовался алгоритм Вильсона метода Ньютона-Рафсона. Были получены следующие оценки параметров $\theta_1 = -0.52$; $\theta_2 = -0.07$. Таким образом, в нашем конкретном случае модель (6) приняла вид

$\theta^{-1}_x(B)x_t = \alpha_t$ с параметрами $\theta_1 = -0.52; \theta_2 = -0.07$, (16)
а модели (7), (8) соответственно

$$\beta_t = \theta^{-1}_x(B)y_t \quad (17)$$

$$\beta_t = v(B)\alpha_t + \xi_t, \text{ где } \xi_t = \theta^{-1}_x(B)n_t. \quad (18)$$

Далее, следуя описанной выше методике, были получены оценки отклика на единичный импульс, анализ поведения которых согласно рекомендациям [17] позволил определить следующие оценки структурных параметров модели передаточной функции: $r = 2, s = 1, b = 3$.

Таким образом, модель передаточной функции принимает вид

$$(1 - \delta_1 B - \delta_2 B^2)y_t = (\omega_0 - \omega_1 B)x_{t-3} + n_t. \quad (19)$$

Начальные оценки «левосторонних» параметров ($\hat{\delta}_i$) можно получить путем решения системы уравнений

$$A\hat{\delta} = h, \quad (20)$$

$$\text{где } a_{ij} = \begin{cases} \hat{v}_{b+s+i-j}, & s+i \geq j \\ 0, & s+i < j \end{cases}; \quad h_i = \hat{v}_{b+s+i}; \quad i, j = 1, 2, \dots, r.$$

Для получения начальных оценок «правосторонних» параметров ($\hat{\omega}_j$) использовались следующие соотношения:

$$\hat{\omega}_0 = \hat{v}_b; \quad (21)$$

если $r \geq s$, то $\hat{\omega}_j = \sum_{i=1}^j \hat{\delta}_i \hat{v}_{b+j-i} - \hat{v}_{b+j}$; если $r < s$, то $\hat{\omega}_j = \sum_{i=1}^j \hat{\delta}_i \hat{v}_{b+j-i} - \hat{v}_{b+j}$ для $j = 1, 2, \dots, r$;

$\hat{\omega}_j = \sum_{i=1}^r \hat{\delta}_i \hat{v}_{b+j-i} - \hat{v}_{b+j}$ для $j = r+1, \dots, s$.

Используя (20), (21), а также полученные ранее оценки структурных параметров модели и оценки отклика на единичный импульс, получаем следующие начальные оценки параметров модели передаточной функции:

$$\hat{\delta}_1 = -0.15; \hat{\delta}_2 = 0.6; \hat{\omega}_0 = -0.08; \hat{\omega}_1 = 0.03.$$

Построение модели шума основывается на восстановлении последовательности n_t путем использования (4) и полученных оценок отклика на единичный импульс, т.е.

$$\hat{n}_t = y_t - \hat{v}_0 \cdot x_t - \hat{v}_1 \cdot x_{t-1} - \dots - \hat{v}_k \cdot x_{t-k},$$

где значение k должно удовлетворять условию $v_{k+i} = 0, i = 1, 2, \dots$. В нашем случае можно взять $k = 14$. К полученной последовательности \hat{n}_t осуществляется подгонка АРСС-модели тем же самым способом, с помощью которого мы строили АРСС-модель для последовательности x_t .

Оценки автокорреляций и частных автокорреляций шума представлены соответственно на рис. 4 и 5.



Рис. 4. Выборочная автокорреляционная функция шума



Рис. 5. Выборочная частная автокорреляционная функция шума

Их анализ позволил определить структуру модели шума в виде (1, 0, 0). После предварительного оценивания модели шума она приняла следующий вид:

$$(1 - 0.53B)n_t = a_t, \quad s_a^2 = 0.78. \quad (22)$$

Уточнение оценок параметров комбинированной модели было выполнено по хорошо известному методу Марквардта [18], позволяющему получить эффективные оценки параметров $b, \bar{\delta}, \bar{\omega}, \bar{\varphi}, \bar{\theta}$ комбинированной модели путем минимизации условной суммы квадратов

$$S_0(b, \bar{\delta}, \bar{\omega}, \bar{\varphi}, \bar{\theta}) = \sum_{u+p+1}^n a_t^2 (b, \bar{\delta}, \bar{\omega}, \bar{\varphi}, \bar{\theta} | x_0, y_0, a_0).$$

После этого комбинированная модель приняла следующий вид:

$$(1 + 0.25B - 0.4B^2)y_t = -0.1x_{t-3} + \frac{1}{1 - 0.77B}a_t \quad (23)$$

или после несложных преобразований

$$(1 - 0.52B - 0.593B^2 + 0.308B^3)y_t = -(0.1 - 0.077B)x_{t-3} + a_t; \quad s_a^2 = 0.73.$$

В целях диагностической проверки модели (23) с использованием (15) было определено значение статистики Q. Оно оказалось равным 15.1 при 13 степенях свободы. Критическое значение этой статистики при указанном числе степеней свободы и уровне значимости 0.05 равно 22.36. Таким образом, можно считать справедливой гипотезу об адекватности разработанной модели.

4. Разработка прогнозного регулятора

Сначала опишем процедуру синтеза регулятора в общем виде для класса моделей

$$Ay(t) = Bx(t-b) + a(t), \quad (24)$$

где

$$A = 1 + a_1S + a_2S^2 + a_3S^3 + \dots + a_{n_a}S^{n_a};$$

$$B = b_0 + b_1S + b_2S^2 + b_3S^3 + \dots + b_{n_b}S^{n_b},$$

S – оператор сдвига назад, т.е. $Sy(t) = y(t-1)$,

а затем применим полученные результаты к модели (23) (легко заметить, что она принадлежит классу моделей (24)). Будем также считать, что горизонты прогнозирования и управления равны между собой ($N_1 = N_2$).

Представим (24) в виде

$$y(t) = \frac{B}{A}x(t-b) + \frac{1}{A}a(t). \quad (25)$$

Тогда уравнение прогноза с упреждением k будет иметь вид

$$y(t+k) = \frac{B}{A}x(t+k-b) + \frac{1}{A}a(t+k). \quad (26)$$

Используя диофантово тождество [15], имеющее для модели (25) вид

$$E_k A = 1 - S^k F_k, \quad (27)$$

где $E_k = e_0 + e_1S + e_2S^2 + \dots + e_{n_e}S^{n_e}$; $F_k = f_0^k + f_1^kS + f_2^kS^2 + \dots + f_{n_f}^kS^{n_f}$;

$k \geq 1$; $n_e = k - 1$; $n_f = n_a - 1$; n_a – порядок полинома A, перепишем уравнение (26) в виде

$$y(t+k) = BE_k x(t+k-b) + F_k y(t) + E_k a(t+k). \quad (28)$$

При этом прогнозное значение выхода с упреждением k можно получить с помощью соотношения

$$\hat{y}(t+k/t) = BE_k x(t+k-b) + F_k y(t). \quad (29)$$

Реальный выход системы может быть записан в виде

$$y(t+k) = \hat{y}(t+k/t) + E_k a(t+k). \quad (30)$$

Полагая, что нужно получить прогнозы для некоторого диапазона значений k (от $k=1$ до $k=N$), запишем уравнение (28) в векторной форме, используя подход, предложенный в [19]:

$$Y = GX + f + a, \quad (31)$$

где

$$\begin{aligned} Y^T &= [y(t+1), y(t+2), \dots, y(t+N)]; \\ X^T &= [x(t-b+1), x(t-b+2), \dots, x(t-b+N)]; \\ f^T &= [F_1 y(t) + d_1, F_2 y(t) + d_2, \dots, F_N y(t) + d_N]; \\ a^T &= [E_1 a(t+1), E_2 a(t+2), \dots, E_N a(t+N)]; \end{aligned} \quad G = \begin{vmatrix} g_0 & 0 & 0 & \dots & 0 \\ g_1 & g_0 & 0 & \dots & 0 \\ g_2 & g_1 & g_0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ g_{N-1} & g_{N-2} & g_{N-3} & \dots & g_0 \end{vmatrix};$$

$$g_i = h_i, \quad i = \overline{0, N-1}; \quad H = BE_N = h_0 + h_1 S^1 + h_2 S^2 + \dots + h_{N-1} S^{(N-1)} + \dots;$$

$$d_k = (BE_k - (h_0 + h_1 S^1 + h_2 S^2 + \dots + h_{k-1} S^{(k-1)}))x(t-b+k), \quad k = \overline{1, N}.$$

Используя векторную форму модели системы (31), запишем критерий оптимальности управления в виде

$$E\{J(t)\} = E\{(GX + f + a - R)^T Q_1 (GX + f + a - R) + (X^T Q_2 X)\} > \min, \quad (32)$$

где $R^T = [r(t+1), r(t+2), \dots, r(t+N)]$ – заданное движение системы; $E\{x\}$ – математическое ожидание x ; Q_1 и Q_2 – диагональные матрицы размерности $N \times N$ с элементами на главных диагоналях соответственно $q_1(1), q_1(2), \dots, q_1(N)$ и $q_2(1), q_2(2), \dots, q_2(N)$.

Дифференцируя (32) по X и приравнявая производную нулю, находим оптимальный вектор X :

$$X = (G^T Q_1 G + Q_2)^{-1} G^T Q_1 (R - f). \quad (33)$$

Выражение (33) определяет оптимальный прогнозный регулятор.

Для вычисления элементов соотношения (33) в работе [20] были предложены полезные формулы, упрощающие расчеты, связанные с синтезом оптимального регулятора.

Коэффициенты полинома E_k (его порядок равен $k-1$) можно получить с помощью соотношения:

$$e_k = -\sum_{i=1}^{n_a} e_{k-i} a_i, \quad k = \overline{1, N-1}, \quad e_0 = 1, \quad e_j = 0 \quad \forall j < 0, \quad (34)$$

где n_a – порядок полинома A ; a_i – его коэффициенты.

Коэффициенты полинома F_k , порядок которого $n_f = n_a - 1$, можно вычислить по формуле

$$f_j^k = -\sum_{i=1}^{n_a} e_{k-i} a_{j+i}, \quad j = \overline{0, n_a - 1}, \quad k = \overline{1, N}, \quad a_n = 0 \quad \forall n > n_a. \quad (35)$$

Параметры матрицы G определяются с помощью следующего соотношения:

$$g_{k-1} = h_{k-1} = b_0 e_{k-1} + b_1 e_{k-2} + \dots + b_{n_b} e_{k-n_b-1}, \quad k = \overline{1, N}, \quad e_j = 0 \quad \forall j < 0. \quad (36)$$

Для вычисления значений d_k целесообразно использовать соотношение:

$$\begin{aligned} d_k &= (b_1 e_{k-1} + b_2 e_{k-2} + \dots + b_{n_b} e_{k-n_b})x(t-b) + \\ &+ (b_2 e_{k-1} + b_3 e_{k-2} + \dots + b_{n_b} e_{k-n_b+1})x(t-b-1) + \\ &\dots \\ &+ b_{n_b} e_{k-1} x(t-b-n_b+1), \quad k = \overline{1, N}, \quad e_j = 0 \quad \forall j < 0. \end{aligned} \quad (37)$$

Рассмотрим синтез прогнозного регулятора для объекта, описываемого моделью (22), при $N=3$. В этом случае имеем:

$$\begin{aligned} n_a &= 3; \quad a_0 = 1; \quad a_1 = -0.52; \quad a_2 = -0.593; \quad a_3 = 0.308; \quad n_b = 1; \\ b_0 &= -1; \quad b_1 = 0.077; \quad n_f = 2; \quad n_e = 2. \end{aligned}$$

Используя описанную выше методику синтеза и взяв в качестве матриц Q_1 и Q_2 диагональные матрицы, все диагональные элементы которых равны $q_1=3.5$ и $q_2=0.02$ соответственно, мы получили следующее представление оптимального регулятора в векторной форме:

$$\begin{bmatrix} x(t) \\ x(t+1) \\ x(t+2) \end{bmatrix} = \begin{bmatrix} -6.030 & 0.403 & -0.915 \\ -1.105 & -6.199 & 0.403 \\ 1.601 & -1.105 & -6.030 \end{bmatrix} \begin{bmatrix} r(t+1) - f(t+1) \\ r(t+2) - f(t+2) \\ r(t+3) - f(t+3) \end{bmatrix},$$

где

$$\begin{aligned} f(t+1) &= 0.520y(t) + 0.593y(t-1) - 0.308y(t-2) + d_1; & d_1 &= 0.077x(t-3); \\ f(t+2) &= 0.863y(t) + 0y(t-1) - 0.160y(t-2) + d_2; & d_2 &= 0.04x(t-3); \\ f(t+3) &= 0.449y(t) + 0.352y(t-1) - 0.266y(t-2) + d_3; & d_3 &= 0.067x(t-3). \end{aligned}$$

Поскольку на практике, как правило, реализуется принцип отступающего горизонта, то особое значение имеет определение управления $x(t)$. После его реализации и получения отклика системы горизонт сдвигается на один такт и управление пересчитывается.

Для $x(t)$ мы имеем

$$\begin{aligned} x(t) &= -6.03(r(t+1) - f(t+1)) + 0.403(r(t+2) - f(t+2)) - 0.915(r(t+3) - f(t+3)) = \\ &= -6.03(r(t+1) - 0.520y(t) - 0.593y(t-1) + 0.308y(t-2) - 0.077x(t-3)) + \\ &+ 0.403(r(t+2) - 0.863y(t) - 0y(t-1) + 0.160y(t-2) - 0.04x(t-3)) - \\ &- 0.915(r(t+3) - 0.449y(t) - 0.352y(t-1) + 0.266y(t-2) - 0.067x(t-3)). \end{aligned} \quad (38)$$

Результаты экспериментального исследования работоспособности разработанного прогнозного регулятора представлены на рис. 6.

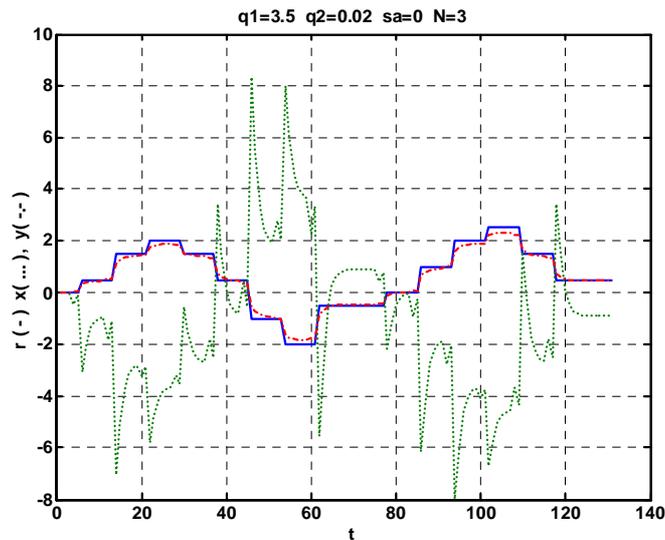


Рис. 6. Результаты моделирования работы регулятора

5. Выводы

1. Впервые синтезирована адекватная ARMAX-модель процесса вытягивания монокристаллических слитков арсенида галлия в условиях промышленной установки «Арсенид-1М», которая может быть использована для регулирования диаметра слитка по каналу температуры фонового нагревателя с целью повысить точность поддержания диаметра.

2. На основе применения разработанной ARMAX-модели процесса выращивания впервые выполнен синтез оптимального прогнозного регулятора для установки «Арсенид-1М», позволяющего учитывать прогнозируемые состояния регулируемого процесса с заданным упреждением и за счет этого обеспечивающего точность поддержания диаметра на уровне ± 2 мм. Эффективность разработанного регулятора подтверждена также приведенными результатами моделирования его работы.

3. Практическая ценность полученных результатов заключается в существенном сокращении потерь дорогостоящего материала при калибровке слитков и повышении качественных показателей выращенных монокристаллов.

4. Дальнейшие исследования следует посвятить разработке математической модели и прогнозного регулятора для процесса выращивания монокристаллических слитков арсенида галлия диаметром 100 мм.

Список литературы: 1. *Satunkin G.A.* Mathematical modelling and control system design of Czochralski and liquid encapsulated Czochralski process: the basic low order mathematical model / G.A. Satunkin // *J. Crystal Growth*. 1995. V. 154. P. 172-188. 2. *Satunkin G. A.* Weighting control of the automatic crystallization process from the melt / G.A. Satunkin, A.G. Leonov // *J. Grystal Growth*. 1990. V. 102. P. 592–608. 3. *Суздаль В.С.* Системы управления процессами получения монокристаллов из расплава / В.С. Суздаль, П.Е. Стадник // *Функциональные материалы для науки и техники. Сборник статей под ред. В.П. Семиноженко*. Харьков: Институт монокристаллов, 2001. С. 514-526. 4. *Петренко В.Р.* Моделирование динамики зависимости диаметра слитка арсенида галлия от скорости его вытягивания / В.Р. Петренко, А.С. Овсепян, Л.Г. Шепель, И.В. Петренко // *Нові технології. Науковий вісник КУЕІТУ*. 2005. №1-2(7-8). С. 46-53. 5. *Оксанич А.П.* Разработка стохастических моделей передаточных функций для системы управления процессом выращивания монокристаллов кремния большого диаметра / А.П. Оксанич, В.Р. Петренко // *Вестник Херсонского государственного технического университета*. 2002. № 2(15). С. 360-363. 6. *Оксанич А.П.* Оценивание адекватности стохастических моделей передаточных функций системы управления процессом выращивания монокристаллов кремния / А.П. Оксанич, В.Р. Петренко // *Нові технології. Науковий вісник Інституту економіки та нових технологій*. 2004. № 3(6). С. 12-14. 7. *Петренко В.Р.* Использование принципа прогнозного управления для регулирования диаметра Cz-Si монокристаллов / В.Р. Петренко, С.В. Ивушкин // *Нові технології. Науковий вісник КУЕІТУ*. 2009. №4(26). С.30-39. 8. *Петренко В.Р.* Регулирование процесса выращивания монокристаллов GaAs по ЛЕС-технологии / В.Р. Петренко, В.А. Тербан // «Автоматизация: проблемы, идеи, решения», міжнародна наук.-технічна конф. (2009; Севастополь). Міжнародна науково-технічна конференція «Автоматизация: проблеми, ідеї, рішення», 7-12 вересня 2009 р.: [матеріали] / редкол.: В.Я. Копп [та ін.]. Севастополь: Вид-во СевНТУ, 2009. С. 281-283. 9. *Model predictive heuristic control* / Richalet J., Rault A., Testud L.J., Papon J. // *Automatica*. 1978. Vol. 14. P. 413-428. 10. *Generalized predictive control. Part 1 and 2.* / Clarke D.W., Mohtadi C., Tuffs P.S. // *Automatica*. 1987. Vol. 23. P. 137-160. 11. *Peterka V.* Predictor-Based Self-Tuning Control / V. Peterka // *Automatica*. 1984. Vol. 20, № 1. P. 39-50. 12. *Properties of Generalized Predictive Control* / D.W. Clarke, C. Mohtadi // *Automatica*. 1989. № 25. P. 859-875. 13. *Analysis and Tuning of Adaptive Generalized Predictive Control* / McIntosh A.R., S.L. Shah, D.G. Fisher // *The Canadian Journal of Chemical Engineering*. 1991. Vol. 69. P. 97-110. 14. *Красовский А.А.* Универсальные алгоритмы оптимального управления непрерывными процессами / А.А. Красовский, В.Н. Буко, В.С. Шендрик. М.: Наука, 1977. 272 с. 15. *Generalized predictive control* / D. Clarke, C. Mohtadi, P. Tu's // *Automatica*. 1987. Vol. 23. P. 137-160. 16. *Allg'ower F., Badgwell T.A., Qin J.S., Rawlings J.B., Wright S.J.* Advances in Control / Highlights of ECC'99 // Chapt. 12. *Nonlinear Predictive Controls and Moving Horizon Estimation*. Springer, London. 1999. P. 391-449. 17. *Анализ временных рядов. Прогноз и управление* / Бокс Дж., Дженкинс Г.; под ред. В.Ф. Писаренко. М.: Мир, 1974. 197 [2]с. 18. *Marquardt D.W.* An Algorithm for least squares estimation of non-linear parameters / D.W. Marquardt // *J. Int. Appl. Math.* 1963. № 11. P. 431-440. 19. *Adaptive general predictive controller for nonlinear systems* / O.M. Zhu, K. Warwick, J.L. Douce // *IEE Proceedings-D*. 1991. Vol. 138, № 1. P. 33-40. 20. *Петренко В.Р.* Синтез оптимального регулятора с предсказанием для процесса выращивания объемных Cz-Si монокристаллов / В.Р. Петренко // *Складні системи і процеси*. 2008. №2(14). С.64-76.

Поступила в редколлегию 11.06.2011

Загирняк Михаил Васильевич, д-р техн. наук, профессор, ректор Кременчугского национального университета им. Михаила Остроградского, зав. кафедрой электрических машин и аппаратов. Научные интересы: автоматизация процессов управления производством полупроводниковых материалов. Адрес: Украина, 39600, Кременчуг, ул. Первомайская, 20, тел.: (05366) 36218. E-mail: mzagirn@kdu.edu.ua.

Оксанич Анатолий Петрович, д-р техн. наук, профессор, директор НИИ технологии полупроводников и информационно-управляющих систем Кременчугского национального университета им. Михаила Остроградского, зав. кафедрой информационно-управляющих систем. Научные интересы: методы и аппаратура контроля структурно-совершенных полупроводниковых монокристаллов. Адрес: Украина, 39600, Кременчуг, ул. Первомайская, 20, тел.: (05366) 30157. E-mail: oksanich@kdu.edu.ua

Петренко Василий Радиславович, д-р техн. наук, профессор кафедры информационно-управляющих систем Кременчугского национального университета им. Михаила Остро-

градского. Научные интересы: автоматизация процессов управления производством полупроводниковых материалов, информационные технологии. Адрес: Украина, 39600, Кременчуг, ул. Первомайская, 20. E-mail: pvt@kdu.edu.ua.

Притчин Сергей Эмильевич, канд. техн. наук, доцент кафедры информационно-управляющих систем Кременчугского национального университета им. Михаила Остроградского. Научные интересы: автоматизация процессов управления производством полупроводниковых материалов. Адрес: Украина, 39600, Кременчуг, ул. Первомайская, 20. E-mail: pritchin@knu.edu.ua

Тербан Виктор Андреевич, канд. техн. наук, доцент, заместитель директора ООО «Силикон». Научные интересы: технологии производства полупроводниковых материалов. Адрес: Украина, 36000, Светловодск, ул. Заводская, 3, тел.: (05236)22041. E-mail: galag@ukrpost.net

**МЕТОДОЛОГИЧЕСКИЕ ПРИНЦИПЫ ПРЕДСТАВЛЕНИЯ АПЕРТУР
ВО МНОЖЕСТВЕ ОДНОМЕРНЫХ ДВУХОСНОВНЫХ
ПОЗИЦИОННЫХ ЧИСЕЛ**

Проводится обоснование и разработка принципов лексикографической нумерации двухосновных позиционных чисел с учетом наличия контекстно-зависимых избыточных последовательностей, которое задает правило нумерации комбинаторных объектов в случае наложения дополнительных запретов, число которых зависит от контекста обработанной последовательности. Излагаются основные этапы построения ассоциативного функционала, задающего правило оценки наличия избыточных одномерных позиционных двухосновных чисел, на основе анализа длины начальной серии нулей в контексте последовательности элементов, предшествующих текущему содержанию.

1. Введение

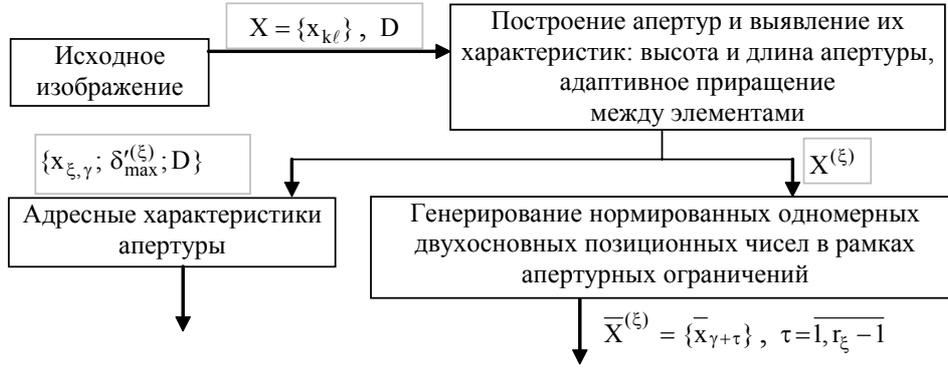
Развитие телекоммуникационных технологий в направлении повышения качества предоставляемых услуг связано с преодолением сложившегося дисбаланса между темпами роста производительности информационных систем и темпами увеличения объемов видеоинформационных потоков [1]. Это диктует *актуальность научно-прикладной тематики исследований*, проводимых в русле сокращения объемов изображений в условиях сохранения их информационного содержания и минимизации сложности реализации [2 – 4]. Данная тематика характерна для вариантов доставки малоразмерных изображений по низкоскоростным беспроводным каналам связи и высококачественных изображений с использованием сравнительно высокоскоростных технологий (10 – 100 Мбит/с). Здесь можно указать такую технологию как WI-FI, основанную на стандарте IEEE 802.11, технологии UWB, 60GHz.

Одно из направлений решения состоит в разработке структурно-комбинаторного представления апертур равномерной длины с полным сохранением их информационного содержания [5]. В то же время для такого подхода обработки апертур отсутствует обоснованная теоретическая база. Поэтому *научная задача* заключается в разработке методологических принципов представления апертур изображений в рамках структурно-комбинаторного подхода на основе двухосновных позиционных чисел. Это позволит в последующем строить эффективные системы кодирования видеопоследовательностей, принадлежащих апертурным ограничениям.

2. Построение источника одномерных двухосновных позиционных чисел

Исходные оцифрованные изображения преобразуются в апертурное описание. В свою очередь последние используются как базовое представление для формирования нормированных одномерных двухосновных позиционных чисел с ограниченным адаптивным приращением между элементами в рамках апертурных ограничений. Это позволяет создать потенциальные возможности для сокращения комбинаторной избыточности без потери качества изображений. Структурно-функциональная схема модели источника генерирова-

ния апертур, рассматриваемых как нормированные одномерные двухосновные позиционные числа (ОДОПЧ), приведена на рис. 1.



Источник нормированных ОДОПЧ

Рис. 1. Структурно-функциональная схема модели источника нормированных ОДОПЧ

На рис. 1. показано, что для исходных последовательностей изображения $X = \{x_{kl}\}$ на основе заданного значения высоты D осуществляется выявление апертур $X^{(\xi)}$. На базе апертур $X^{(\xi)}$, элементы которых удовлетворяют условию $x_{\xi, \gamma+\tau} \in [\ell_{\xi}^{(\min)}; \ell_{\xi}^{(\max)}]$, $\tau = \overline{0, r_{\xi} - 1}$, проводится формирование нормированных одномерных двухосновных позиционных чисел $\bar{X} = \{\bar{x}_{\gamma+\tau}\}$.

3. Разработка методологических принципов описания апертур двухосновными позиционными числами

Для непосредственного сокращения избыточности сообщений, образуемых источником нормированных ОДОПЧ, требуется построить систему формирования кодовых конструкций. При создании системы кодирования необходимо учитывать, что:

1. ОДОПЧ являются последовательностями, принадлежащими множеству $\Psi(\delta_{\max}^{(\xi)}; D)$, формируемому с учетом условий, которые задаются следующими выражениями:

– для обобщенных ограничений на элементы аперттуры

$$\begin{cases} \ell_{\xi}^{(\min)} = x_{\xi, \gamma} - D/2 \leq x_{\xi, \gamma+\tau} - \delta \leq x_{\xi, \gamma+\tau+1} \leq x_{\xi, \gamma} + D/2 = \ell_{\xi}^{(\max)}; \\ \ell_{\xi}^{(\max)} = x_{\xi, \gamma} + D/2 \geq x_{\xi, \gamma+\tau} + \delta \geq x_{\xi, \gamma+\tau+1} \geq x_{\xi, \gamma} - D/2 = \ell_{\xi}^{(\min)}; \end{cases}$$

$\tau = \overline{0, r_{\xi} - 1}$, где $x_{\xi, \gamma}$ – вершина аперттуры; r_{ξ} , D – соответственно длина и высота аперттуры; δ – приращение между элементами;

– для динамических диапазонов элементов $x_{\xi, \gamma+\tau}$ относительно предыдущего элемента аперттуры $\psi_{\xi, \gamma+\tau} = 2\delta + 1$, для $\tau = \overline{1, r_{\xi} - 1}$;

– для условия выхода минимальных и максимальных абсолютных значений элементов ОДОПЧ за границы аперттуры $x_{\eta, \min} < x_{\xi, \gamma} - D/2$ и $x_{\eta, \max} > x_{\xi, \gamma} + D/2$.

2. Объем $W(\delta_{\max}^{(\xi)}; D)$ множества $\Psi(\delta_{\max}^{(\xi)}; D)$ определяется по формуле

$$\begin{aligned} W(\delta_{\max}^{(\xi)}; D) &= (D+1)(2\delta_{\max}^{(\xi)} + 1)^{r_{\xi} - 1} - (1 - \text{sign}(1 + \text{sign}(\frac{D}{2} - \eta_{\text{exc}} \delta_{\max}^{(\xi)}))) \times \\ &\times 2 \sum_{\tau = \eta_{\text{exc}}}^{r_{\xi} - 1} \left((\eta_{\text{exc}} \delta_{\max}^{(\xi)} - \frac{D}{2})(2\delta_{\max}^{(\xi)} + 1)^{r_{\xi} - \tau - 1} \right). \end{aligned}$$

3. Требуется обеспечить взаимоднозначное соответствие между элементами ОДОПЧ и кодовой конструкцией.

С учетом первого условия предлагается формировать кодовое представление ОДОПЧ как номер соответствующей последовательности во множестве $\Psi(\delta_{\max}^{(\xi)}; D)$. При этом для того, чтобы обеспечить сокращение требуемого вида комбинаторной избыточности, необходимо сформировать систему кодирования (нумерации), зависящую от вершины апертуры, ее высоты, величины адаптивного приращения между элементами и учитывающую наличие количества исключаемых контекстно-зависящих ОДОПЧ, а также проведение нормировки относительно вершины апертуры.

Следующей теоремой обоснуем принцип лексикографической нумерации в условиях наличия избыточных последовательностей, количество которых зависит от контекста обработанных элементов.

Теорема 1 (о правиле лексикографической нумерации). *Правило лексикографической нумерации нормированного ОДОПЧ $\bar{X}^{(\xi)} = \{\bar{x}_{\gamma+\tau}\}$, $\tau = \overline{1, r_{\xi} - 1}$ с учетом исключения количества контекстно-зависимых избыточных последовательностей определяется формулой (без учета вершины апертуры)*

$$W'(\bar{X}^{(\xi)}) = C(\delta_{\max}^{(\xi)})_{r_{\xi}-1} = \sum_{\tau=1}^{r_{\xi}-1} \sum_{\ell=x_{\min}}^{\bar{x}_{\xi, \gamma+\tau}-1} \left(W(\bar{x}_{\xi, \gamma+1}, \bar{x}_{\xi, \gamma+2}, \dots, \bar{x}_{\xi, \gamma+\tau-1}, \ell) - \varphi_c(\bar{x}_{\xi, \gamma+1}, \bar{x}_{\xi, \gamma+2}, \dots, \bar{x}_{\xi, \gamma+\tau-1}) \Delta W(\bar{x}_{\xi, \gamma+1}, \bar{x}_{\xi, \gamma+2}, \dots, \bar{x}_{\xi, \gamma+\tau-1}, \ell) \right), \quad (1)$$

где $C(\delta_{\max}^{(\xi)})_{r_{\xi}-1} - W'(\bar{X}^{(\xi)})$ – номер последовательности $\bar{X}^{(\xi)}$ в упорядоченном множестве допустимых ОДОП чисел; $W(\bar{x}_{\xi, \gamma+1}, \bar{x}_{\xi, \gamma+2}, \dots, \bar{x}_{\xi, \gamma+\tau-1}, \ell)$ – количество одномерных двухосновных позиционных чисел, у которых первые τ элементов равны $(\bar{x}_{\xi, \gamma+1}, \bar{x}_{\xi, \gamma+2}, \dots, \bar{x}_{\xi, \gamma+\tau-1}, \ell)$, без учета исключения контекстно-запрещенных последовательностей; $\varphi_c(\bar{x}_{\xi, \gamma+1}, \bar{x}_{\xi, \gamma+2}, \dots, \bar{x}_{\xi, \gamma+\tau-1})$ – функционал, задающий правило определения наличия избыточных последовательностей, на основе анализа контекста $(\bar{x}_{\xi, \gamma+1}, \bar{x}_{\xi, \gamma+2}, \dots, \bar{x}_{\xi, \gamma+\tau-1})$ предшествующих элементов; $\Delta W(\bar{x}_{\xi, \gamma+1}, \bar{x}_{\xi, \gamma+2}, \dots, \bar{x}_{\xi, \gamma+\tau-1}, \ell)$ – количество контекстно-запрещенных последовательностей для подмножества ОДОПЧ, у которых первые τ элементов равны $(\bar{x}_{\xi, \gamma+1}, \bar{x}_{\xi, \gamma+2}, \dots, \bar{x}_{\xi, \gamma+\tau-1}, \ell)$.

Доказательство. Условимся, что упорядочивание последовательностей осуществляется по наибольшему значению старшего элемента в случае равенства их предыдущих элементов. Обозначим количество одномерных двухосновных позиционных чисел, у которых первые τ элементов равны $(\bar{x}_{\xi, \gamma+1}, \bar{x}_{\xi, \gamma+2}, \dots, \bar{x}_{\xi, \gamma+\tau-1}, \bar{x}_{\xi, \gamma+\tau})$, без учета исключения контекстно-запрещенных последовательностей как $W(\bar{x}_{\xi, \gamma+1}, \bar{x}_{\xi, \gamma+2}, \dots, \bar{x}_{\xi, \gamma+\tau-1}, \bar{x}_{\xi, \gamma+\tau})$. Тогда без учета сокращения контекстно-запрещенных последовательностей код-номер $C_{r_{\xi}-1}$ искомой последовательности на основе принципа упорядочивания определяется по формуле

$$C_{r_{\xi}-1} = \sum_{\tau=1}^{r_{\xi}-1} \sum_{\ell=x_{\min}}^{\bar{x}_{\xi, \gamma+\tau}-1} \left(W(\bar{x}_{\xi, \gamma+1}, \bar{x}_{\xi, \gamma+2}, \dots, \bar{x}_{\xi, \gamma+\tau-1}, \ell) \right).$$

Предложим, что введенный принцип упорядочивания выполняется и для избыточных последовательностей. Отсюда для подсчета количества избыточных последовательностей введем величину $\Delta W(\bar{x}_{\xi, \gamma+1}, \bar{x}_{\xi, \gamma+2}, \dots, \bar{x}_{\xi, \gamma+\tau-1}, \bar{x}_{\xi, \gamma+\tau})$, определяемую как количество контекстно-запрещенных последовательностей для подмножества ОДОПЧ, у которых первые τ элементов равны $(\bar{x}_{\xi, \gamma+1}, \bar{x}_{\xi, \gamma+2}, \dots, \bar{x}_{\xi, \gamma+\tau-1}, \bar{x}_{\xi, \gamma+\tau})$. Поскольку количество контекстно-запрещенных ОДОПЧ и их появление зависит от последовательности предыдущих элементов (контекста), то введем функционал $\varphi_c(\bar{x}_{\xi, \gamma+1}, \bar{x}_{\xi, \gamma+2}, \dots, \bar{x}_{\xi, \gamma+\tau-1})$. Данный функционал формализует правило определения наличия избыточных последовательностей на основе анализа контекста $(\bar{x}_{\xi, \gamma+1}, \bar{x}_{\xi, \gamma+2}, \dots, \bar{x}_{\xi, \gamma+\tau-1})$ предшествующих элементов относи-

тельно τ -го элемента. В соответствии с определением допустимых нормированных ОДОПЧ их количество формируется с учетом устранения контекстно-запрещенных последовательностей. Тогда искомое значение $C(\delta'_{\max}^{(\xi)})_{\tau-1}$ кода-номера последовательности $\bar{X}^{(\xi)}$ в упорядоченном множестве допустимых ОДОПЧ вычисляется по формуле (1). Теорема 1 доказана.

Выражение (1) задает правило лексикографической нумерации множества комбинаторных объектов в случае наложения дополнительных запретов, число которых зависит от текущего контекста обрабатываемой последовательности.

Следствие из теоремы 1. Количество $W'(\bar{x}_{\xi, \gamma+1}, \bar{x}_{\xi, \gamma+2}, \dots, \bar{x}_{\xi, \gamma+\tau-1}, \bar{x}_{\xi, \gamma+\tau})$ допустимых последовательностей для каждого τ -го элемента обрабатываемой последовательности

определяется как разность между $\sum_{\chi=1}^{\tau} \sum_{\ell=x_{\min}}^{\bar{x}_{\xi, \gamma+\chi-1}} (W(\bar{x}_{\xi, \gamma+1}, \bar{x}_{\xi, \gamma+2}, \dots, \bar{x}_{\xi, \gamma+\chi-1}, \ell))$ накопленным количеством ОДОПЧ без учета контекстно-избыточных последовательностей и $\sum_{\chi=1}^{\tau} \sum_{\ell=x_{\min}}^{\bar{x}_{\xi, \gamma+\chi-1}} (\Delta W(\bar{x}_{\xi, \gamma+1}, \bar{x}_{\xi, \gamma+2}, \dots, \bar{x}_{\xi, \gamma+\chi-1}, \ell))$ накопленным количеством контекстно-запрещенных последовательностей, которые являются избыточными.

Функционал $\Phi_{c, \tau}$ в соотношении (1) должен обеспечить в формализованном виде проведение анализа контекста, который состоит из элементов, предшествующих обрабатываемому. Структура функционала определяется правилами формирования апертур фиксированной длины, на базе которых образуются контекстно-запрещенные последовательности. Для получения функционала $\Phi_{c, \tau}$ сформулируем и докажем следующую теорему.

Теорема 2 (о формализации правила проверки наличия контекстно-запрещенных последовательностей). Функционал $\Phi_{c, \tau-1}$, обеспечивающий определение наличия контекстно-запрещенных ОДОПЧ имеет следующий вид:

$$\Phi_{c, \tau-1} = \Phi_c(x_{\xi, \gamma} = 0, \bar{x}_{\xi, \gamma+1}, \bar{x}_{\xi, \gamma+2}, \dots, \bar{x}_{\xi, \gamma+\tau-1}) = (1 - \text{sign}(\tau - v_{0, \tau-1})), \quad (2)$$

где $v_{0, \tau-1}$ – длина начальной серии нулей в значениях признаков направлений приращений; $(1 - \text{sign}(\tau - v_{0, \tau-1}))$ – решающее правило, указывающее на наличие избыточных последовательностей, у которых соответствующие элементы предшествуют текущему τ -му элементу ОДОПЧ:

$$1 - \text{sign}(\tau - v_{0, \tau-1}) = \begin{cases} 1, & \rightarrow v_{0, \tau-1} = \tau; \\ 0, & \rightarrow v_{0, \tau-1} < \tau. \end{cases} \quad (3)$$

Доказательство. В соответствии с правилами формирования апертур фиксированной длины для τ -го элемента одномерного двухосновного позиционного числа будет существовать множество контекстно-запрещенных последовательностей, если:

1) существует позиция η_{exc} , для которой происходит выход значений элементов ОДОПЧ за границы апертуры (рис. 2);

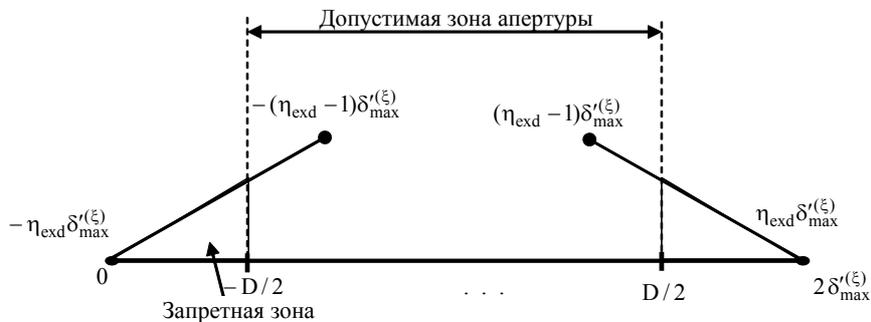


Рис. 2. Схема выхода возможных элементов ОДОПЧ за границы апертуры

2) значения первых τ элементов обрабатываемой последовательности равны нулю, т.е. $\{x_{\xi, \gamma+v}=0\}$, $v=0, (\tau-1)$.

Для оценки выполнения второго условия предлагается использовать величину $v_{0, \tau-1}$, равную длине непрерывной серии нулей, выявленной для последовательности, которая составлена из значений признаков выполнения условия $\{x_{\xi, \gamma+v}=0\}$, начиная с вершины апертуры $x_{\xi, \gamma}$. Если условие $\{x_{\xi, \gamma+v}=0\}$ выполняется, то значение соответствующего v -го признака будет равно нулю. И наоборот, когда $\{x_{\xi, \gamma+v} \neq 0\}$, то значение признака будет отличным от нулевого значения. Для вычисления признака условия $\{x_{\xi, \gamma+v}=0\}$ используется параметр u - индекс зоны нормирования, определяемый как:

$$u = \text{sign}(\bar{x}_{\xi, \gamma+\tau} - \bar{x}_{\xi, \gamma+\tau-1} + \delta'_{\max}(\xi) - \delta'_{\max}(\xi)) = \text{sign}(\bar{x}_{\xi, \gamma+\tau} - \bar{x}_{\xi, \gamma+\tau-1}) = \begin{cases} -1, & \rightarrow \bar{x}_{\xi, \gamma+\tau} < \bar{x}_{\xi, \gamma+\tau-1}; \\ 0, & \rightarrow \bar{x}_{\xi, \gamma+\tau-1} = \bar{x}_{\xi, \gamma+\tau}; \\ 1, & \rightarrow \bar{x}_{\xi, \gamma+\tau} > \bar{x}_{\xi, \gamma+\tau-1}. \end{cases} \quad (4)$$

Понятно, что условие $\{x_{\xi, \gamma+v}=0\}$ выполняется, если для соответствующих значений элементов апертуры будет выполняться равенство $u = 0$. Под серией нулей понимается последовательность подряд идущих нулевых значений. В случае, когда появляется единичное значение, то текущая серия считается оконченной. Это позволяет выявлять только те варианты, для которых условие $\{x_{\xi, \gamma+v}=0\}$ выполняется для всех анализируемых элементов. Для учета такой зависимости вводится функция $1 - \text{sign}(\tau - v_{0, \tau-1})$, которая будет принимать единичное значение только в случае, когда длина нулевой серии признаков u будет равной количеству предшествующих элементов, т.е. $v_{0, \tau-1} = \tau$. Отсюда значение функционала φ_c для τ -го элемента ОДОПЧ будет равно единице только тогда, когда условие $\{x_{\xi, \gamma+v}=0\}$ выполняется для всех $v=0, (\tau-1)$. Теорема 2 доказана.

Из теоремы 2 вытекает, что соотношения (2) – (4) образуют систему решающих правил, на основе которых определяется наличие контекстно-запрещенных последовательностей относительно обрабатываемого элемента одномерного двухосновного позиционного числа, т.е. наличие базовых элементов $x_{\xi, \eta_{\text{exc}}+\tau}$ так, что $x_{\xi, \eta_{\text{exc}}+\tau} \in [-\eta_{\text{exc}} \delta'_{\max}(\xi); -(D/2) - 1]$.

В случае, когда $\varphi_c(x_{\xi, \gamma} = 0, x_{\xi, \gamma+1}, x_{\xi, \gamma+2}, \dots, x_{\xi, \gamma+\tau-1}) = 0$, то это означает, что для обрабатываемого ОДОПЧ отсутствуют неучтенные контекстно-запрещенные (КЗ) последовательности.

Следующей особенностью обработки ОДОПЧ является нормированность их элементов. В результате нормировки элементов $x_{\xi, \gamma+\tau}$ апертуры относительно координаты вершины апертуры $x_{\xi, \gamma}$ формируются последовательности \bar{X}_{ξ} (нормированные ОДОПЧ), значения элементов которых $\bar{x}_{\xi, \gamma+\tau}$ изменяются в диапазоне:

$$\bar{x}_{\xi, \gamma+\tau-1} - \delta'_{\max}(\xi) \leq \bar{x}_{\xi, \gamma+\tau} \leq \bar{x}_{\xi, \gamma+\tau-1} + \delta'_{\max}(\xi), \quad \tau = \bar{1}, r_{\xi} - 1. \quad (5)$$

Как следует из полученного соотношения, разница между элементами ОДОПЧ ограничена величиной $\delta'_{\max}(\xi)$ и может принимать как положительные, так и отрицательные значения. В то же время согласно принятому лексикографическому правилу элементы обрабатываемых последовательностей должны принимать положительные значения с минимумом, равным нулю. Поэтому для сокращения динамического диапазона обрабатываемых элементов и перевода их в область положительных значений предлагается перейти к дифференциальному представлению. Тогда формируются компоненты $(\bar{x}_{\xi, \gamma+\tau} - \bar{x}_{\xi, \gamma+\tau-1})$,

Список литературы: **1.** *Олифер В.Г.* Компьютерные сети. Принципы, технологии, протоколы: Учебник для вузов. 3-е изд. / В.Г. Олифер, Н.А. Олифер. СПб.: Питер, 2006. 958 с. **2.** *Gonzales R.C.* Digital image processing / R.C. Gonzales, R.E. Woods. Prentice Inc. Upper Saddle River, New Jersey. 2002. 779 p. **3.** *Миано Дж.* Форматы и алгоритмы сжатия изображений в действии: Учебное пособие / Дж. Миано; пер. с англ. М. : Триумф, 2003. 336 с. **4.** *Сэломон Д.* Сжатие данных, изображений и звука. М: Техносфера, 2004. 368 с. **5.** *Баранник В.В.* Структурно-комбинаторное представление данных в АСУ / В.В. Баранник, Ю.В. Стасев, Н.А. Королева. Х.: ХУПС, 2009. 252 с.

Поступила в редколлегию 19.10.2011

Баранник Владимир Викторович, д-р техн. наук, профессор, ведущий научный сотрудник научного центра Харьковского университета Воздушных Сил им. Ивана Кожедуба. Научные интересы: системы, технологии преобразования, кодирования, защиты и передачи информации, семантической обработки изображений. Адрес: Украина, 61023, Харьков, ул. Сумская, 77/79, тел. 050-3038971.

Кальченко Дмитрий Сергеевич, соискатель ХНУРЭ. Научные интересы: системы, технологии преобразования, кодирования, защиты и передачи информации. Адрес: Украина, 61166, Харьков, пр. Ленина, 14.

УДК 658.52.011.56

С.А.КАНЦЕДАЛ, М.В.КОСТИКОВА

ОДИН АЛГОРИТМ РЕШЕНИЯ ЗАДАЧИ О ПОКРЫТИИ

Для решения задач о покрытии среднего размера предлагается простой алгоритм, реализующий ограниченный перебор сочетаний покрывающих объектов. Приводится способ вычислений предельного количества основных операций, выполняемых алгоритмом.

1. Постановка задачи

Рассмотрим самую общую постановку задачи о покрытии. Для конечного множества $M = \{m_1, m_2, \dots, m_m\}$ и некоторого конечного семейства его подмножеств $S_j, j=1, 2, \dots, n$ требуется найти подсемейство $S'_j \subseteq S_j$ с минимальным набором подмножеств S_j такое, что каждый элемент исходного множества M принадлежит, по крайней мере, одному из этих подмножеств. Найденные подмножества S'_j обычно называют покрывающими, а семейство S'_j – наименьшим покрытием M .

В том случае, когда покрывающим элементам приписаны некоторые веса, рассматривают взвешенную задачу о покрытии, предполагающую поиск наименьшего взвешенного покрытия M . Когда же покрывающие подмножества попарно не пересекаются, задача о наименьшем покрытии трактуется как задача о разбиении множества M на наименьшее количество классов.

Цель исследования – разработка нового алгоритма задачи о покрытии.

Задача для достижения цели – обзор существующих методов решения такой задачи.

Приведем простейший пример взвешенной задачи о покрытии. Некоторой организации требуется нанять переводчиков с ряда языков на английский. Имеется пять кандидатур таких переводчиков A, B, C, D, E , делающих переводы только с части языков и требующих свою оплату труда.

Необходимо определить, каких переводчиков следует взять на работу, чтобы обеспечить возможность перевода со всех языков на английский и при этом гарантировать наименьшие расходы организации.

Исходные данные задачи представлены двумя таблицами (табл. 1, 2), в первой из которых в соответствующих ее клетках возможности переводчиков, т.е. перевода ими с приведенного перечня языков на английский, отмечены единицами, невозможности – нулями. Очевидно, что множество M , для которого требуется найти наименьшее взвешенное покрытие, образует перечень языков. Кандидатуры переводчиков представляют собой семейство подмножеств $S_j = \{A, B, C, D, E\}$, подсемейство $S'_j \subseteq S_j$ – перечень переводчи-

ков, обеспечивающих возможность перевода, т.е. покрытия M с минимальными затратами.

По существу первая таблица без верхней строки и крайнего левого столбца представляет собой нуль-единичную матрицу. Строки матрицы интерпретируются как языки, столбцы – как переводчики. Поэтому задачу можно толковать как поиск такого множества столбцов матрицы, что в целом они обеспечивают наличие единиц во всех ее строках, образно говоря, покрывают строки этой матрицы и приводят к наименьшим расходам.

Таблица 1. Возможность переводчика переводить на английский язык с указанного языка

Язык \ Переводчики	A	B	C	D	E
Французский	1	0	1	1	0
Немецкий	0	1	0	0	1
Греческий	0	1	0	0	0
Итальянский	1	0	0	1	0
Испанский	0	0	1	0	0
Русский	0	1	1	0	1
Китайский	1	0	0	1	1

Таблица 2. Оплата труда переводчиков

Переводчики	A	B	C	D	E
Оплата труда	35000	40000	38000	45000	47000

Другим примером является задача о развозке грузов. Она моделируется так, что в целом граф G представляет сеть дорог. Одна из вершин графа представляет склад, а остальные его вершины – торговые точки (магазины, кафе и т.п.). Транспортное средство (автомобиль), покидающее склад, снабжает товарами некоторые из этих точек, получающие весь необходимый ассортимент. После этого автомобиль возвращается на склад.

Заранее с учетом практических ограничений проезда выбрано некоторое число маршрутов $1, 2, \dots, j, \dots, n$ со стоимостями $C_1, C_2, \dots, C_j, \dots, C_n$.

Требуется определить, сколько машин следует использовать на маршрутах, чтобы в течение дня доставить все грузы со склада потребителям и обеспечить наименьшую суммарную стоимость всех замкнутых маршрутов.

В этой задаче исходное множество M – это перечень торговых точек, семейство подмножеств S_j – перечень маршрутов, $S'_j \subseteq S_j$ – подсемейство маршрутов, обеспечивающих минимальное покрытие M .

Как и выше, получаем нуль-единичную матрицу, строки которой $1, 2, \dots, i, \dots, m$ интерпретируются как торговые точки, а столбцы – как маршруты. Наличие единицы на пересечении некоторой строки и столбца в этой матрице означает, что данная торговая точка включена в соответствующий маршрут. Наличие нуля отображает тот факт, что она не входит в рассматриваемый маршрут.

Таким образом, задача сводится к тому, чтобы в указанной матрице найти то количество столбцов, которые покрывают строки матрицы и минимизируют расходы на перевозки.

Чисто формально задача о наименьшем покрытии легко представима в терминах целочисленного линейного программирования. Исходим из того, что задана нуль-единичная матрица $P = \|P_{ij}\|$, строки которой $i = 1, 2, \dots, m$ – это индексы элементов исходного множества M , столбцы $j = 1, 2, \dots, n$ – индексы элементов семейства S , а элементы P_{ij} равны единице или нулю в зависимости от того, покрывает j -й столбец i -ю строку или нет. Введем булевы переменные $x_j, j = 1, 2, \dots, n$ такие, что $x_j = 1$, если j -й столбец в искомым покрытии входит в него и $x_j = 0$ в противном случае. С переменными x_j свяжем веса

c_j (в рассматриваемом примере оплаты переводчиков). В результате получаем задачу: минимизировать функцию

$$Z = \sum_{j=1}^n c_j x_j$$

при условиях

$$P_{ij} = \begin{cases} 1, & \text{если } j\text{-й столбец покрывает } i\text{-ю строку,} \\ 0 & \text{в противном случае,} \end{cases}$$

$$x_j = \begin{cases} 1, & \text{если } j\text{-й столбец входит в покрытие,} \\ 0 & \text{в противном случае,} \end{cases}$$

$$\sum_{j=1}^n P_{ij} x_j \geq 1, \quad i = 1, 2, \dots, m.$$

При этом последнее неравенство гарантирует покрытие строк матрицы P элементами x_j .

2. Метод решения задачи

В связи с тем, что задача целочисленного линейного программирования NP-полная, задача о наименьшем покрытии также обладает этим свойством, т.е. точно может быть решена только алгоритмами экспоненциального характера.

Первыми методами решения задачи были методы отсечения Гомори, среди которых наилучшим оказался усовершенствованный его алгоритм – алгоритм Мартина [1]. Позже для решения задачи были предложены несколько модификаций метода ветвей и границ [2, 3].

Для задач относительно небольшого размера ($m \leq 30$, $n \leq 20$) мы предлагаем простейший алгоритм ее решения – последовательной генерации и проверки сочетаний. Его действия продемонстрируем на задаче набора переводчиков.

На первом шаге сформируем все варианты пар переводчиков (A, B) , (A, C) , (A, D) , (A, E) , (B, C) , (B, D) , (B, E) , (C, D) , (C, E) , (D, E) и для каждой пары установим возможность покрытия строк исходной матрицы. Если такая возможность существует больше, чем для одной пары, выбираем ту из них, которая определит наименьшие затраты и завершим поиск. В противном случае, т.е. когда покрытия парами нет, сформируем последовательность троек переводчиков (A, B, C) , (A, B, D) , (A, B, E) , (A, C, D) , (A, C, E) , (A, D, E) , (B, C, D) , (B, C, E) , (B, D, E) , (C, D, E) , определим возможность покрытия, в зависимости от чего завершим решение задачи или продолжим формирование состава переводчиков. Описанные действия продолжим до обнаружения первого покрывающего их сочетания. Если такого сочетания нет, задача не имеет решения. Заметим, что проверка возможности покрытия выполняется достаточно просто. В каждой строке матрицы суммируются элементы столбцов, образующих сочетание, и при обнаружении первой нулевой суммы в некоторой ее строке делается вывод о том, что данное сочетание не обеспечивает покрытие. В противном случае покрытие существует.

Применительно к данной задаче убеждаемся, что покрытие обеспечивает две тройки переводчиков: (A, B, C) и (B, C, D) . Для первой тройки расходы равны $35000 + 40000 + 38000 = 113000$ денежных единиц, для второй – $40000 + 38000 + 45000 = 123000$ таких единиц. Следовательно, искомым решением является прием на работу переводчиков A , B , C .

3. Оценка трудоемкости метода

Оценим сверху трудоемкость алгоритма, который легко может быть построен путем реализации описанных действий. Для этого подсчитаем количество основных операций – сложений и сравнений, – которые он будет выполнять для установления того факта, имеет ли задача решение или нет.

В связи с тем, что число вариантов двоек, троек, $(n-1)$ -к столбцов матрицы P равно числу сочетаний C_n^2 , C_n^3 , C_n^{n-1} , всего для выбора получим $C_n^2 + C_n^3 + \dots + C_n^{n-1}$ вариантов

набора ее столбцов. В свою очередь для проверки любого варианта на покрытие в худшем случае требуется выполнить m сравнений и некоторое количество сложений, зависящее от числа элементов в сочетании. Так, для вариантов, образуемых парами столбцов матрицы P , для каждой ее строки необходимо выполнить одно сложение и одно сравнение для выявления нуля, т.е. в целом для матрицы $2m$ операций. Варианты, образуемые тройками столбцов, для каждой строки матрицы P требуют выполнить два сложения и одно сравнение, т.е. всего $3m$ операций. Варианты с $(n-1)$ столбцами требуют выполнить $n-2$ сложения и одно сравнение, т.е. $(n-1)$ -ую операцию, в каждой строке P , а в целом для матрицы P $-(n-1)m$ операцию. Наконец, последний вариант с n столбцами в каждой строке P требует выполнить $n-1$ сложение и одно сравнение, т.е. в целом для матрицы $n \cdot m$ операцию.

Таким образом, предельное количество основных операций, выраженное через число сочетаний, следующее $\left[2C_n^2 + 3C_n^3 + \dots + (n-1)C_n^{n-1} + n\right]m$. Окончательно для получения верхней оценки числа операций, выполняемых алгоритмом, необходимо элементы приведенного выражения умножить на число операций, которое требуется выполнить для генерации одного сочетания. Согласно [4], где приводится два алгоритма построения множеств сочетаний, генерация одного сочетания потребует то число операций, которое не больше, чем значение верхнего индекса символа сочетания операций. В результате окончательно для верхней оценки получим $W = \left[4C_n^2 + 9C_n^3 + \dots + (n-1)^2 C_n^{n-1} + n\right]m$.

Безусловно, полученная оценка очень завышена, так как определяет то число операций, которое нужно выполнить для установления факта, существует ли решение или нет. В действительности число операций должно быть существенно меньше, поскольку предполагается, что часть задач будут иметь решения. К сожалению, найти аналитическое выражение для вычисления такого числа операций не представляется возможным, так как нельзя предсказать, сколько столбцов матрицы P определит решение. Единственная возможность установить среднее и максимальное число этих операций, а также их рост в зависимости от размера задачи – набор компьютерной статистики, ее обработка и аппроксимация полученных данных подходящей эмпирической кривой. Прикидочный расчет числа операций по приведенной формуле для задачи $m = 20, n = 10$ даст

$$W = \left(4 \cdot \frac{10!}{2!8!} + 9 \cdot \frac{10!}{3!7!} + 16 \cdot \frac{10!}{4!6!} + 25 \cdot \frac{10!}{5!5!} + 36 \cdot \frac{10!}{6!4!} + 49 \cdot \frac{10!}{7!3!} + 64 \cdot \frac{10!}{8!2!} + 81 \cdot \frac{10!}{9!} + 10\right) \times \\ \times 20 = (180 + 1080 + 3360 + 6300 + 7560 + 5880 + 2880 + 810 + 10) \cdot 20 = 28060 \cdot 20 = 561200$$

операций, что вполне приемлемо для решения задач среднего размера на современных компьютерах.

В заключение решим предложенным алгоритмом задачу о наименьшем взвешенном покрытии, исходные данные (P, C) которой приведены ниже.

Составим сочетания для всех пар столбцов матрицы P и, если пара покрывает ее строки, отметим ее знаком плюс (+), в противном случае – знаком минус (-). Имеем $(1,2)^-, (1,3)^-, (1,4)^-, (1,5)^-, (1,6)^-, (2,3)^-, (2,4)^-, (2,5)^-, (2,6)^-, (3,4)^-, (3,5)^-, (3,6)^-, (4,5)^-, (4,6)^-, (5,6)^-$. В связи с тем, что пары индексов столбцов матрицы P не образуют покрытия, переходим к формированию сочетаний из трех индексов P . В результате получаем $(1,2,3)^-, (1,2,4)^-, (1,2,5)^-, (1,2,6)^-, (1,3,4)^-, (1,3,5)^-, (1,3,6)^-, (1,4,5)^+, (1,4,6)^-, (1,5,6)^-, (2,3,4)^-, (2,3,5)^-, (2,3,6)^-, (2,4,5)^-, (2,4,6)^+, (2,5,6)^-, (3,4,5)^-, (3,4,6)^-, (3,5,6)^-, (4,5,6)^-$.

	1	2	3	4	5	6
1	1	1	1	0	0	1
2	1	0	1	0	0	1
$P = 3$	0	0	0	1	0	0
4	0	1	0	0	1	0
5	0	0	0	0	1	1
6	1	1	0	0	0	0
$C =$	4	7	5	8	3	2

Таким образом, покрытие дают сочетания столбцов 1, 4, 5 со стоимостью $C_1 + C_4 + C_5 = 4 + 8 + 3 = 15$ и 2, 4, 6 со стоимостью $C_2 + C_4 + C_6 = 7 + 8 + 2 = 17$, откуда оптимальное решение задачи – столбцы 1, 4, 5 матрицы P и стоимость 15. Как видим, это решение достигается при его выборе из $C_6^2 + C_6^3$ сочетаний, на что затрачивается $W = (4 \cdot C_6^2 + 9 \cdot C_6^3 + 6) \cdot 6$, т.е. не более, чем 1476 основных операций.

Выводы

Научная новизна работы – получен новый алгоритм решения задачи о покрытии для задач среднего размера.

Практическая значимость – алгоритм легко программируем, в связи с чем он может быть без затруднений использован в производственных целях.

Список литературы: 1. Корбут А.А., Финкельштейн Ю.Ю. Дискретное программирование. М.: Наука, 1969. 368 с. 2. Кристофидес Н. Теория графов. Алгоритмический подход. М.: Мир, 1978. 432 с. 3. Lemke C.E., Salkin H.M., Spielberg K. Set covering by single branch enumeration with linear programming Subproblems // Oper. Res. 1971. V. 19. P. 998 – 1015. 4. Лунский В. Комбинаторика для программистов. М.: Мир, 1988. 213 с.

Поступила в редколлегию 04.06.2011

Канцедал Сергей Андреевич, д-р техн. наук, профессор, зав. кафедрой финансы и кредит Западнодонецкого института экономики и управления. Научные интересы: математическое моделирование, теория расписаний и её применение. Адрес: Украина, Павлоград, ул. Днепровская, 400.

Костикова Марина Владимировна, канд. техн. наук, доцент кафедры информатики Харьковского национального автомобильно-дорожного университета. Научные интересы: математическое моделирование, теория расписаний и её применение. Адрес: Украина, 61002, Харьков, ул. Петровского, 25, тел. 707-37-74.

МОДЕЛЬ КОМБІНОВАНОЇ ІНФРАСТРУКТУРИ ВІДКРИТИХ КЛЮЧІВ

Пропонується модель комбінованої інфраструктури відкритих ключів (ІВК), що поєднує у собі переваги ІВК на базі X.509 та ІВК на базі ідентифікаторів. Особливістю розробленої комбінованої ІВК є наявність відокремленого серверу підпису, який забезпечує цілісність, автентичність, доступність та своєчасне поновлення бази даних ідентифікаторів користувачів домену та загальносистемних параметрів.

Вступ

Як показують дослідження, подальший розвиток електронного документообігу можливий лише за умови суттєвого скорочення витрат на ІВК [1]. На даний момент в Україні ІВК функціонують на національному або регіональному рівні та базуються на використанні сертифікатів відповідно до стандарту X.509 [2]. Використання ІВК на базі X.509 у межах організацій, міністерств та відомств не дозволить скоротити витрати користувачів, оскільки побудова та розгортання такої інфраструктури вимагає значних фінансових затрат. З іншого боку, використання альтернативної ІВК на базі ідентифікаторів дозволяє скоротити витрати на створення, впровадження та користування послугами, що надає зазначена інфраструктура [3]. Однак архітектура ІВК на базі ідентифікаторів потребує довіри до уповноваженого на генерації ключів (УГК), оскільки безпосередньо УГК генерує та видає особистий ключ користувачу. Вимога розповсюдження та можливість зберігання особистих ключів користувачів не дозволяє використовувати ІВК на базі ідентифікаторів на регіональному та національному рівнях.

У науковому співтоваристві з'явилася ідея використання комбінованої ІВК, тобто ІВК на базі X.509 впроваджується на національному та регіональному рівнях, а ІВК на базі ідентифікаторів – на рівні організацій, міністерств та відомств [4].

На даний час запропоновано декілька моделей комбінованих ІВК, що мають за мету спростити процес електронного документообігу за рахунок скорочення кількості сертифікатів користувачів. Розглянемо основні недоліки запропонованих моделей комбінованих ІВК.

Girault [5] будує схему, у якій відкритий ключ ділиться на дві частки, одну з яких виробляє сервер, а іншу – користувач. У такому разі УГК не в змозі підмінити відкритий ключ користувача. Але для того щоб довести справжність відкритого ключа, відправнику необхідно виконати інтерактивний протокол з нульовими знаннями з одержувачем. Зазначена схема передбачає, що відправник та одержувач володіють однаковими загальносистемними параметрами.

У схемі Gentry [6] таємний ключ розділяється на дві частини, однією з них володіє сервер, іншою – користувач. Для розшифрування повідомлення користувач повинен отримати другу компоненту таємного ключа від сервера по захищеному каналу зв'язку. Сервер сертифікує ідентифікатор та відкритий ключ користувача. Таким чином, усі користувачі повинні мати сертифікати.

Al-Riyami та Paterson [7] запропонували схему, дуже схожу на схему Gentry, але у ній немає необхідності сертифікувати відкритий ключ користувача. Користувач публікує дві частки відкритого ключа, які пов'язані між собою. Але сервер генерації ключів може легко скомпрометувати користувача, замінюючи його відкритий ключ, та при цьому ця заміна не може бути ніяким чином виявлена.

Chen у роботі [8] пропонує механізм перевірки відкритих ключів засвідчувальних центрів, що видають особисті ключі. Він показує, яким чином можна побудувати ієрархічну систему, у якій можна перевірити увесь ланцюг сертифікатів. Але поза увагою залишаються питання отримання справжніх ідентифікаторів та загальносистемних параметрів.

Callas [9] запропонував схему, у якій сервер генерує відкрий і таємний ключ кожного користувача. Щоб отримати відкритий ключ користувача, необхідно здійснити неавтентифікований запит до сервера. Для отримання таємного ключа користувача необхідно здійснити автентифікований запит до сервера. Механізми доведення цілісності та справжності відкритого та таємного ключів не уточнюються.

У схемі Voltage [3] користувачі володіють сертифікатом традиційної ІВК та ключем ІВК на ідентифікаторах. Сертифікати використовуються для автентифікації користувача та для накладання цифрового підпису, ключі ІВК на ідентифікаторах — для направленої шифрування. Головною перевагою системи, на думку розробників, є відсутність перевірки сертифікату одержувача перед відправкою повідомлення. Механізми взаємодії користувачів з різних доменів та дії у разі компрометації ключів ІВК на ідентифікаторах не уточнюються.

Зазначимо, що у всіх цих системах не вирішується проблема призначення та розповсюдження відкритих ідентифікаторів. Очевидно, що у інформаційній системі, яка налічує мільйони користувачів, практично неможливо розробити політику призначення загальновідомих та відкритих ідентифікаторів.

Ціль – розробка моделі комбінованої ІВК, що має переваги ІВК як на базі X.509, так і на базі ідентифікаторів; розробка методу направленої шифрування, що дозволяє вирішити проблемні питання взаємодії користувачів та узгодження загальносистемних параметрів у такій ІВК.

Задачі: 1) розробка моделі комбінованої інфраструктури відкритих ключів; 2) розробка протоколу узгодження загальносистемних параметрів та взаємодії користувачів.

Об'єкт дослідження – процеси криптографічних перетворень у ІВК при реалізації узгодження загальносистемних параметрів та направленої шифрування.

Предмет дослідження – модель комбінованої ІВК, метод направленої шифрування, що дозволяють налагодити взаємодію користувачів з доменів, які застосовують різні загальносистемні параметри.

Попередні дослідження показують, що створювати ІВК на ідентифікаторах має сенс тільки у рамках деякої організації. Таким чином, будуть існувати багато доменів з різними системними параметрами та політиками безпеки. Серед проблемних питань, які необхідно вирішити для побудови такої системи, є питання взаємодії користувачів різних доменів та узгодження загальносистемних параметрів цих доменів.

1. Архітектура комбінованої ІВК

Сутність розробленої комбінованої інфраструктури відкритих ключів полягає у тому, що уся множина користувачів поділяється на групи (домени). На практиці групою можна вважати деякий відділ, організацію, міністерство тощо. Усі користувачі певної групи використовують ІВК на базі ідентифікаторів. Таким чином, структура групи щонайменше складається з УГК та користувачів. Зазначимо, що УГК кожної групи визначає загальносистемні параметри, політику безпеки та ідентифікатори користувачів, які мають бути цілісними, автентичними та доступними у межах групи. Ідентифікатори користувачів можуть бути представлені у вигляді бази даних.

Можливість перевірки користувачем будь-якої групи цілісності та автентичності бази даних ідентифікаторів, загальносистемних параметрів та політики безпеки довільно фіксованої групи досягається шляхом використання ІВК на базі X.509. Зазначимо, що у роботі архітектура ІВК на базі X.509 не визначається у повному обсязі. Вважається, що використовуються усі необхідні сервіси, які надає традиційна ІВК (OCSP, TSP, реєстрація користувачів тощо).

Особливістю розробленої комбінованої ІВК є наявність серверу підпису, який є елементом групи та забезпечує цілісність, автентичність, доступність та своєчасне поновлення бази даних ідентифікаторів користувачів групи шляхом накладання електронного цифрового підпису (ЕЦП). Сервер підпису використовує пару ключів, що визначена ІВК на базі X.509, тобто сервер підпису має сертифікат відкритого ключа. Цілісність та автентичність загальносистемних параметрів домену і політики безпеки досягається за рахунок додання їх до розширень сертифікату відкритого ключа серверу підпису. Таким чином, будь-який користувач може перевірити цілісність та справжність загальносистемних параметрів та політики безпеки шляхом перевірки ланцюжка сертифікатів. Зауважимо, що користувачі групи не є суб'єктами ІВК на базі X.509, але програмне забезпечення, яке вони використовують, підтримує необхідні алгоритми перевірки ЕЦП.

Схематично модель комбінованої ІВК зображена на рис. 1. На схемі у якості УГК виступає сервер ключів.

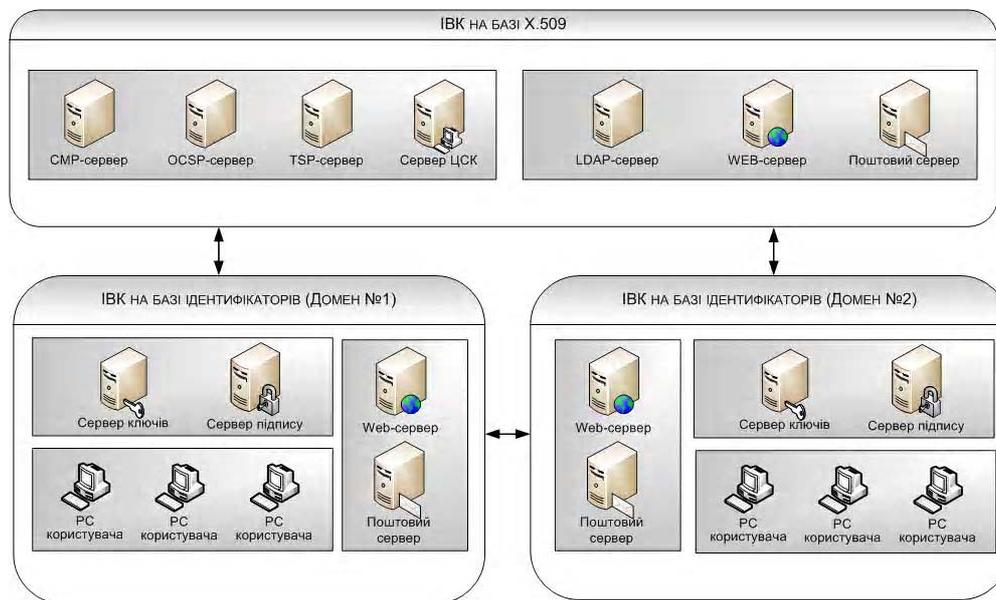


Рис. 1. Архітектура комбінованої інфраструктури відкритих ключів

Розглянемо механізм направленої шифрування (НШ). У випадку належності користувачів одному домену схема направленої шифрування відбувається тільки у межах ІВК на базі ідентифікаторів, наприклад з використанням схеми Boneh-Franklin [10]. Після формування зашифрованого повідомлення користувач відправляє його іншому користувачу через поштовий сервер. Одержувач повідомлення отримує особистий ключ з серверу ключів або, у разі наявності ключа, одразу розшифровує повідомлення. У такому випадку поштовий сервер визначає відправника повідомлення для одержувача.

Якщо користувачі належать різним доменам, то відправник направлено зашифровує повідомлення та підписує його, використовуючи алгоритми НШ, що визначені у зовнішній ІВК на ідентифікаторах, та ЕЦП, що визначені у внутрішній ІВК на ідентифікаторах відповідно. Далі сервер підпису отримує підписане та зашифроване повідомлення, перевіряє ЕЦП, тобто переконується, що зашифроване повідомлення цілісне та дійсно належить користувачу групи. У разі успішної перевірки сервер підпису видаляє ЕЦП користувача, підписує тільки зашифроване повідомлення та відправляє його одержувачу. Значимо, що сервер підпису не спроможний розшифрувати повідомлення відправника. Наступним кроком поштовий сервер у групі одержувача самостійно або за допомогою серверу підпису перевіряє ЕЦП та у разі успіху відправляє одержувачу. Одержувач повідомлення отримує особистий ключ з серверу ключів, або у разі наявності ключа одразу розшифровує повідомлення.

Слід відзначити, що схема НШ є досить повільною для великих обсягів даних, тому доцільно використовувати алгоритм симетричного шифрування. Наприклад, відправник генерує ключ та зашифровує повідомлення відповідно до ГОСТ 28147[11], а наступним кроком направлено зашифровує згенерований ключ з використанням схеми Boneh-Franklin [10].

2. Протоколи взаємодії користувачів комбінованої ІВК

Розглянемо формалізовану модель комбінованої інфраструктури ІВК. Нехай задана множина доменів $\Omega = \{\Omega_i \mid i = 1..m\}$, де $m > 0$ - кількість доменів. Вважатимемо, що домен Ω_i складається з серверу генерації ключів S_{KG}^i , серверу підписів S_{DS}^i та множини користувачів $U_i = \{u_{ij} \mid j = 1..n_i\}$, тобто $\Omega_i = \{S_{KG}^i, S_{DS}^i, U_i\}$.

Позначимо центр сертифікації ключів (ЦСК) у ІВК на базі X.509 як S_{PKI} .

Визначимо протокол $\Sigma = \{I, R, U, V\}$ взаємодії користувачів у комбінованій інфраструктурі ключів, як протокол, що складається з чотирьох етапів:

- Етап ініціалізації I .
- Етап генерації ключів користувачів R .
- Етап направлено шифрування U .
- Етап розшифрування V .

Розглянемо **етап ініціалізації**, що складається з встановлення системних параметрів S_{PKI} та з встановлення системних параметрів доменів Ω_i .

Етап встановлення системних параметрів S_{PKI} . Обирається просте число q , група G та генератор групи P порядку q . Центр сертифікації ключів генерує майстер-ключ $d \in [1, q-1]$ та обчислює відкритий ключ $Q = dP$. Позначимо множину загальносистемних параметрів як $D = \{q, G, P\}$. Зазначимо, що відкритий ключ Q та множина загальносистемних параметрів D визначені в сертифікаті відкритого ключа S_{PKI} .

Етап встановлення системних параметрів Ω_i . S_{DS}^i генерує особистий ключ $d_{DS}^i \in [1, q-1]$ та обчислює відкритий ключ $Q_{DS}^i = d_{DS}^i P$. Далі вибирається просте число q_i , дві групи G_1^i, G_2^i порядку q_i та білінійне відображення Вейля або Тейта $e_i : G_1^i \times G_1^i \rightarrow G_2^i$ та генератор групи $P_i \in G_1^i$. S_{KG}^i генерує майстер-ключ $d_{KG}^i \in Z_{q_i}^*$ та обчислює відкритий ключ $Q_{KG}^i = d_{KG}^i P_i$. Далі обираються криптографічні геш-функції $H_1^i : \{0,1\}^* \rightarrow G_1^i$ та $H_2^i : G_2^i \rightarrow \{0,1\}^l$ для деякого l . Таким чином, d_{DS}^i, Q_{DS}^i - це пара ключів традиційної ІВК, а d_{KG}^i, Q_{KG}^i - пара ключів ІВК на базі ідентифікаторів. Системними параметрами домену Ω_i будуть $D_i = \{q_i, G_1^i, G_2^i, \hat{e}_i, P_i, H_1^i, H_2^i\}$. Відкритий ключ Q_{DS}^i та множина загальносистемних параметрів D_i визначені в сертифікаті відкритого ключа S_{DS}^i . Позначимо базу ідентифікаторів користувачів домену Ω_i як $DB_i = \{ID_{ij} \mid j = 1..n_i\}$. База даних DB_i підписується на ключі d_{DS}^i .

Етап генерації ключів користувачів визначається так:

Кожному користувачу u_{ij} домену Ω_i відповідає ідентифікатор $ID_{ij} \in \{0,1\}^*$. Будь-якому ідентифікатору $ID_{ij} \in \{0,1\}^*$ однозначно відповідає відкритий ключ $Q_{ij} = H_1^i(ID_{ij}) \in G_1^i$ та таємний ключ $d_{ij} = d_{KG}^i Q_{ij}$, що формується по запити користувача до S_{KG}^i .

Процес встановлення ключових даних має такий вигляд (рис. 2).

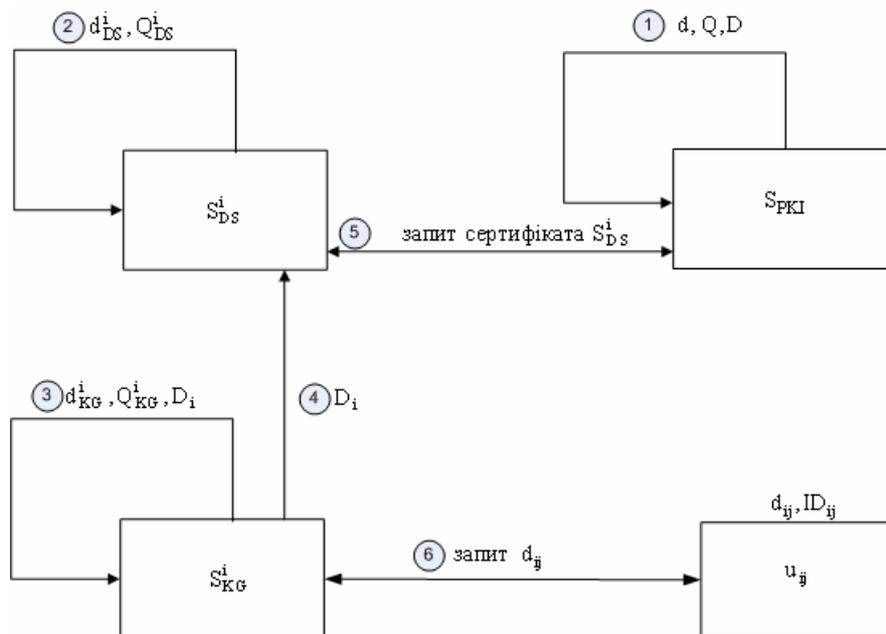


Рис. 2. Схема розподілення ключових даних

Наступним кроком визначимо протоколи направлено шифрування та розшифрування у межах різних доменів (рис. 3) або одного і того ж домену (рис. 4). Зафіксуємо домени Ω_i та Ω_r . Оберемо двох користувачів $u_{ij} \in \Omega_i$ та $u_{rs} \in \Omega_r$ і розглянемо два випадки $i = r$ (один домен) та $i \neq r$ (різні домени).

Визначимо **етап направлено шифрування**.

У випадку $i \neq r$ протокол направлено шифрування користувача u_{ij} на користувача u_{rs} описується так:

1. Користувач u_{ij} звертається до веб-серверу домену Ω_r та отримує сертифікат відкритого ключа серверу підпису S_{DS}^r та базу DB_r . Наступним кроком користувач u_{ij} перевіряє справжність сертифікату, перевіряючи підпис S_{PKI} за допомогою відкритого ключа Q , та отримує з нього загальносистемні параметри D_r . Далі користувач u_{ij} перевіряє цілісність бази даних шляхом перевірки цифрового підпису на відкритому ключі Q_{DS}^r та отримує відкритий ідентифікатор ID_{rs} користувача u_{rs} .

2. Користувач u_{ij} направлено зашифрує повідомлення M , використовуючи відкритий ключ Q_{rs} та загальносистемні параметри D_r , а потім підписує його на своєму особистому ключі d_{ij} . У результаті отримується зашифроване та підписане повідомлення $S_{d_{ij}}(E_{Q_{rs}}(M))$, яке він надсилає на сервер цифрового підпису S_{DS}^i .

3. Сервер S_{DS}^i перевіряє підпис користувача u_{ij} та у разі успіху підписує зашифроване повідомлення $E_{Q_{rs}}(M)$ на особистому ключі d_i . Отримане повідомлення $S_{d_i}(E_{Q_{rs}}(M))$ надсилається на поштовий сервер домену Ω_r .

Аналогічним чином визначимо **етап розшифрування** для випадку $i \neq r$.

1. Поштовий сервер домену Ω_r звертається до веб-серверу домену Ω_i , та отримує сертифікат серверу S_{DS}^i . Наступним кроком поштовий сервер домену Ω_r перевіряє справжність сертифікату. Поштовий сервер домену Ω_r перевіряє цілісність повідомлення шляхом перевірки цифрового підпису на відкритому ключі Q_{DS}^i . У разі успіху поштовий сервер домену Ω_r надсилає повідомлення користувачу u_{rs} .

2. Користувач u_{rs} , у разі відсутності у нього особистого ключа, звертається до сервера ключів S_{KG}^r з запитом на отримання ключа.

3. Сервер ключів генерує d_{rs} та надсилає його користувачу u_{rs} .

4. Користувач u_{rs} розшифровує повідомлення M за допомогою ключа d_{rs} .

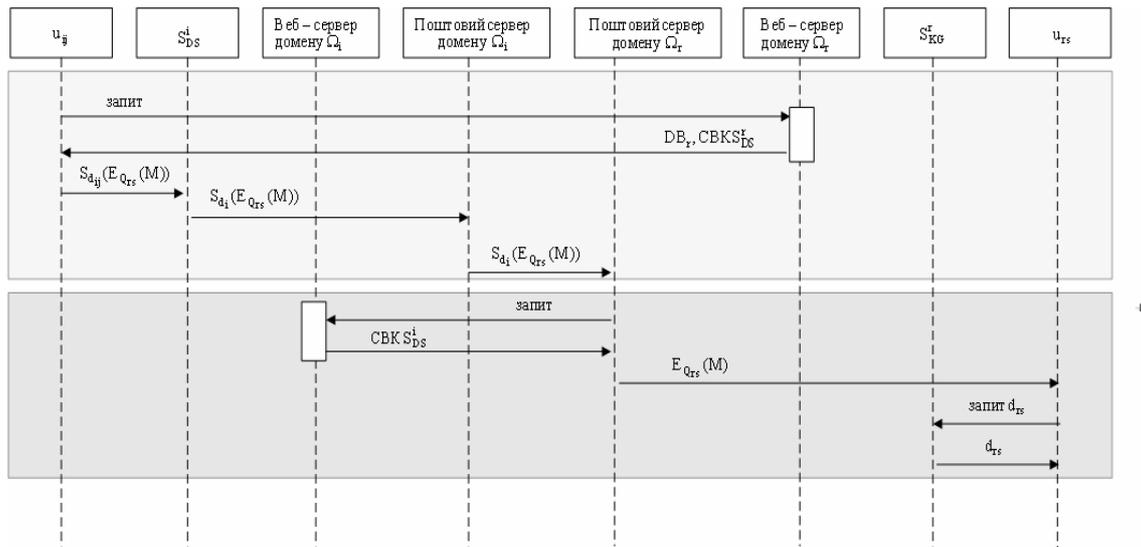


Рис. 3. Протокол взаємодії користувачів

У випадку $i = r$ **етап направлено шифрування** користувача u_{ij} на користувача u_{is} має вигляд.

1. Користувач u_{ij} звертається до веб-серверу домену Ω_i та отримує сертифікат відкритого ключа серверу підпису S_{DS}^i та базу DB_i . Наступним кроком користувач u_{ij} перевіряє справжність сертифікату, перевіряючи підпис S_{PKI} за допомогою відкритого ключа Q , та отримує з нього загальносистемні параметри D_i . Далі користувач u_{ij} перевіряє цілісність бази даних шляхом перевірки цифрового підпису на відкритому ключі Q_{DS}^i та отримує відкритий ідентифікатор ID_{is} користувача u_{is} .

2. Користувач u_{ij} направлено зашифрує повідомлення M , використовуючи відкритий ключ Q_{is} та загальносистемні параметри D_i , і надсилає користувачу u_{is} через поштовий сервер домену Ω_i .

Етап розшифрування для випадку $i = r$.

1. Користувач u_{is} , у разі відсутності у нього особистого ключа, звертається до сервера ключів S_{KG}^i з запитом на отримання ключа.

2. Сервер ключів генерує для користувача u_{is} особистий ключ та надсилає його користувачу u_{is} .

3. Користувач u_{is} розшифровує повідомлення M за допомогою ключа d_{is} .

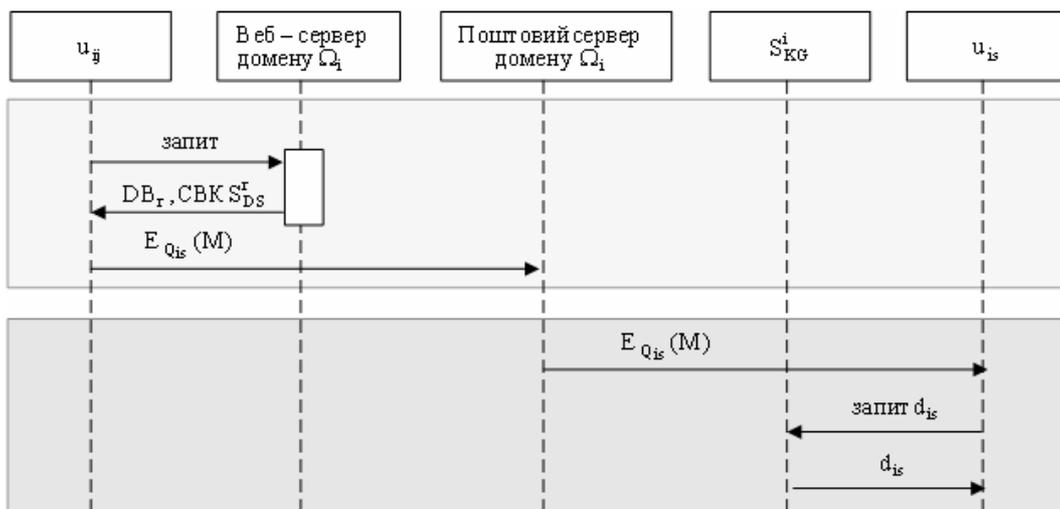


Рис. 4. Протокол взаємодії користувачів

3. Вирішення проблемних питань при розгортанні та використанні комбінованої ІВК

Щоб практично побудувати комбіновану ІВК, яка дозволить знизити вартість послуг безпеки, необхідно вирішити такі питання:

1. Розповсюдження загальносистемних параметрів.
2. Розроблення політики безпеки.

Розглянемо більш детально проблему розповсюдження загальносистемних параметрів. По-перше, зауважимо, що криптографічні примітиви у ІВК на базі X.509 та на базі ідентифікаторів не можуть використовувати однакові загальносистемні параметри. Криптографічні примітиви у ІВК на базі ідентифікаторів на поточний момент базуються на використанні еліптичних кривих, що не задовільняють MOV-умові, у той час як криптографічні примітиви у ІВК на базі X.509 використовують еліптичні криві, що задовільняють зазначеній умові. Що стосується множини доменів, у яких функціонує ІВК на базі ідентифікаторів, то у загальному випадку слід вважати, що для кожного домену задані свої загальносистемні параметри.

Таким чином, постає задача розповсюдження автентичних загальносистемних параметрів з метою унеможливлення атак, що базуються на підміні відкритих параметрів системи. Загальносистемні параметри ІВК на базі X.509 містять в сертифікаті, тобто їх цілісність та справжність гарантується ЦСК, що випустив сертифікат. Щодо загальносистемних параметрів ІВК на базі ідентифікаторів, то дослідниками було запропоновано декілька підходів [4]. Параметри можуть бути включені до сертифікату відкритого ключа як додаткові поля, але у разі компрометації особистого ключа, що відповідає відкритому ключу сертифіката, доведеться перевипустити сертифікат. Альтернативою цьому підходу є використання сертифікату атрибутів для загальносистемних параметрів. Але зважаючи на те, що в Україні не підтримуються сертифікати атрибутів, у запропонованій комбінованій ІВК використовується перший підхід.

Сервер генерації ключів визначає загальносистемні параметри домену та передає їх серверу підпису, який додає їх до свого запиту на сертифікат. Фактично, якщо сервер генерації передав загальносистемні параметри серверу підпису, то він повинен звернутися із запитом до ЦСК на перевипуск сертифікату. Але загальносистемні параметри не змінюються часто, тому це суттєво не вплине на частоту перевипуску сертифікату. У випадку блокування або скасування сертифікату серверу підпису загальносистемні параметри немає потреби змінювати, а необхідно лише додати до нового сертифікату відкритого ключа серверу підпису.

При розробці політики безпеки домену слід вирішити питання формування та розповсюдження ідентифікаторів користувачів домену, а також порядок генерації та видачі особистих ключів користувачів домену.

Відкриті ідентифікатори користувачів повинні бути загальнодоступними та зручними для запам'ятовування строками. Пропонується використовувати внутрішні поштові адреси. Одна із проблем, яку необхідно вирішити, була описана у роботі Callas [5]. Її сутність полягає у тому, що у разі відкритого доступу до усіх ідентифікаторів групи зловмисник може дослідити її структуру або розсилати спам. Щоб запобігти цьому, пропонується використовувати відкриту базу ідентифікаторів (яка містить ідентифікатори публічних співробітників організації) та закрити, до якої мають доступ тільки співробітники.

Проблема захищеного отримання ключа від центру сертифікації є однією з найважливіших для інфраструктури відкритих ключів. Традиційна ІВК вирішує це питання за допомогою отримання таємного ключа безпосередньо у приміщенні центру сертифікації. Такий підхід може бути прийнятним у разі довгострокових ключів, якщо ж ключі є короткостроковими, то отримання особистого ключа у приміщенні сторонньої організації призводить до значних накладних витрат. У схемі компанії Voltage це питання вирішується шляхом суворої автентифікації користувача на ключовому сервері з використанням сертифікатів. Таким чином, користувач повинен володіти довгостроковим ключем для отримання короткострокового, що суперечить ідеї ІВК на базі ідентифікаторів, згідно з якою користувачі не мають сертифікатів. У запропонованій моделі ключовий сервер розташовується у межах тієї ж групи, що і користувач, тобто витрати часу на отримання особистого ключа будуть значно менші. Також пропонується використовувати неінтерактивний метод. Необхідно зазначити, що на відміну від традиційної ІВК, користувач не повинен отримувати особистий ключ на кожний період, а тільки тоді, коли у цьому є потреба.

Відкриті ключі змінюються автоматично за рахунок конкатенації відкритого ідентифікатора користувача та мітки часу (період зміни якої встановлюється у політиці безпеки відповідної групи).

До переваг розробленої системи можна віднести:

1. Масштабованість системи.
2. Використання ідентифікаторів як відкритих ключів для кінцевих користувачів.
3. Можливість взаємодії користувачів з різних доменів з різними загальносистемними параметрами.
4. Забезпечення конфіденційності, цілісності та справжності.
5. Можливість підключення додаткових сервісів (антивірус, антиспам тощо).
6. Можливість контролю кореспонденції з боку керівництва (якщо є потреба).

Основними недоліками розробленої системи є:

1. Навантаження на сервер підпису домену (слід виконувати формування та перевірку підпису для кожного електронного листа).
2. Необхідність перепідписувати базу ідентифікаторів при реєстрації або видаленні нового користувача.
3. Розробка апаратних модулів, що підтримували б криптографічні примітиви у ІВК на базі ідентифікаторів.

4. Висновки

Наукова новизна представлена новою моделлю комбінованої ІВК, яка поєднує ІВК на базі X.509 та ІВК на базі ідентифікаторів. Особливістю розробленої моделі комбінованої ІВК є наявність відокремленого серверу підпису, який забезпечує цілісність, автентичність, доступність та своєчасне поновлення бази даних ідентифікаторів користувачів домену та загальносистемних параметрів.

Також був розроблений метод направленої шифрування у комбінованій ІВК для взаємодії користувачів з різних доменів, що мають різні загальносистемні параметри та політики безпеки.

Практична значущість полягає у зменшенні вартості усієї системи, що обумовлено скороченням кількості сертифікатів завдяки застосуванню ІВК на базі ідентифікаторів на рівні кінцевих користувачів. Практична значимість методу направленої шифрування полягає у можливості взаємодії користувачів доменів з різними загальносистемними параметрами. Сутність методу полягає у формуванні підписаних загальносистемних параметрів, отриманні та перевірці параметрів користувачем іншого домену, направленому шифруванні

повідомлення на отриманих параметрах та перевірці цілісності отриманого зашифрованого повідомлення сервером підпису.

Зазначимо, що основними недоліками системи є навантаження на сервер підпису та необхідність перепідписування бази ідентифікаторів при реєстрації або видаленні нового користувача. Проте домен комбінованої ІВК налічує в середньому не більш ніж 500 користувачів, тому загальне навантаження на сервер підпису не є великим.

У подальшому планується визначити структури розширень сертифікатів, що відносяться до ІВК на базі ідентифікаторів, а також формати підписаних та зашифрованих даних.

Список літератури: **1.** Горбенко І.Д., Онопрієнко В.В. Стан та проблемні питання створення та розвитку Національної ІВК // Прикладна радіоелектроніка. 2006. Том.5. №1. С. 41-51 **2.** ITU-T (International Telecommunications Union) Recommendation X.509: Information Technology - Open Systems Interconnection The Directory: Authentication Framework. 2000 **3.** *Voltage Security. Identity-Based Encryption and PKI Making Security Work.* 2005. **4.** Geraint Price, Chris J. Mitchell: Interoperation Between a Conventional PKI and an ID-Based Infrastructure. EuroPKI 2005: 73-85. **5.** Girault M. Self-certified public keys. In D. W. Davies, editor, Advances in Cryptology—EUROCRYPT'91, volume 547 of Lecture Notes in Computer Science. 1992. P. 490–497. Springer-Verlag. **6.** Gentry C. Certificate-based encryption and the certificate revocation problem. In E. Biham, editor, Advances in Cryptology – EUROCRYPT 2003, volume 2656 of Lecture Notes in Computer Science, pages 272–293. Springer-Verlag, 2003. **7.** Al-Riyami S., Paterson K. G. Certificateless Public Key Cryptography, extended abstract in Proceedings of ASIACRYPT '03, LNCS 2894, Springer-Verlag, 2003. **8.** Chen L., Harrison K., Moss A., Soldera D., and Smart N.P. Certification of public keys within an identity based system. In A. H. Chan and V. D. Gligor, editors, Information Security, 5th International Conference, ISC, volume 2433 of LNCS, pages 322–333. Springer-Verlag, 2002. **9.** Callas J. Identity-Based Encryption with Conventional Public-Key Infrastructure. In 4th Annual PKI R&D Workshop, number 7224 in Interagency Reports, pages 102–115. NIST, 2005. **10.** Boneh D. and Franklin M. Identity-based encryption from the Weil pairing. Advances in Cryptology - CRYPTO 2001, volume 2139 of Lecture Notes in Computer Science. Springer-Verlag. 2001. P. 213-229. **11.** ГОСТ 28147-89 “Государственный стандарт Союза ССР. Система обработки информации. Защита криптографическая. Алгоритм криптографического преобразования”.

Поступила в редколлегию 01.06.2011

Погребняк Костянтин Анатолійович, канд. техн. наук, доцент каф. БІТ ХНУРЕ. Адреса: Україна, 61166, Харків, пр. Леніна, 14, e-mail: iitkostya@gmail.com

Кравченко Павло Олександрович, аспірант каф. БІТ ХНУРЕ. Адреса: Україна, 61166, Харків, пр. Леніна, 14, e-mail: kravchenko@gmail.com

УДК 681.325.53:004.5

*Ю.В. ЛОПУХИН, Е.В. БОЧАРОВ, В.В. ВАРЕЦА, К.В. ПОЛЕЖАЕВ,
А.Н. МАКАРЕНКО*

СТАТИСТИЧЕСКИЙ АНАЛИЗ ПРЕОБРАЗОВАНИЙ ЧИСЕЛ В ПРЕОБРАЗОВАТЕЛЯХ КОДОВ ПАРАЛЛЕЛЬНОГО ТИПА

Рассматривается структура и возможности программного средства "TRANSFORMATION" для определения основных статистических характеристик преобразователей кодов, использующих метод накопления эквивалентов и параллельную стратегию.

1. Постановка задачи

Преимуществами преобразователей кодов по методу накопления эквивалентов (ПК НКЭ) являются умеренные аппаратные затраты и возможность изменения в широких пределах соотношения между числом тактов преобразования (быстродействием) и аппаратными затратами за счет выбора числа шагов преобразования, их оптимальных значений, числа блоков разбиения многоразрядного ПК НКЭ и стратегии использования шагов преобразования (последовательной или параллельной).

При применении последовательной стратегии в многоразрядных ПК НКЭ преобразование чисел ведется путем уменьшения значений преобразуемых цифр (в тех разрядах, показания, где значение равно или превышает значение шага) на величину большего шага.

При этом формируется соответствующий эквивалент, который на каждом такте преобразования добавляется к содержимому накапливающего сумматора.

Если же во всех преобразуемых разрядах значения цифр оказываются меньше большего шага, то происходит переход на более меньший шаг и т. д. вплоть до использования шага 1 и обнуления всех разрядных счетчиков. Значения шагов преобразования должны выполняться следующие ограничения:

$$1 < a < b; \quad a < b < c; \quad b < c \quad K-1, \quad (1)$$

где K – основание системы счисления на входе; a, b, c – соответственно второй, третий и четвертый шаги преобразования (первый шаг преобразования всегда равен 1).

Последовательная стратегия по сравнению с параллельной структурно реализуется достаточно просто. Оценки сверху для числа тактов преобразования ПК НКЭ являются несколько завышенными, что приводит к необходимости разработки программного средства, позволяющего найти не оценки сверху для числа тактов преобразования, а математическое ожидание M и дисперсию D числа тактов преобразования.

Применение параллельной стратегии преобразования чисел в ПК НКЭ приводит к уменьшению числа тактов преобразования (увеличению быстродействия) за счет того, что в различных преобразуемых разрядах разрешается одновременное вычитание различных шагов.

Параллельная стратегии применима только в ПК НКЭ с числом шагов преобразования, равным двум и более. Для нахождения основных статистических характеристик параллельной стратегии (математического ожидания и дисперсии) и их сравнения для последовательной и параллельной стратегий разработано программное средство "TRANSFORMATION".

Целью настоящей работы является:

- рассмотрение назначения и возможностей разработанного программного средства;
- анализ функционирования программного средства в режиме Статистика;
- сравнительная оценка статистических характеристик последовательной и параллельной стратегий преобразований чисел.

2. Параллельная стратегия преобразования чисел в ПК НКЭ

Стратегия преобразования чисел в ПК НКЭ (как было отмечено ранее) может быть последовательной или параллельной. При последовательной стратегии быстродействие ПК НКЭ можно повысить за счет увеличения числа шагов с двух до трех и более. При этом относительный эффект увеличения быстродействия по сравнению с 2-шаговым ПК НКЭ снижается, а аппаратные затраты существенно возрастают. Более выгодной является параллельная стратегия преобразования чисел, при которой допускается одновременное вычитание различных шагов в различных преобразуемых разрядах (табл. 1).

Как следует из табл.1, количество тактов преобразования числа (в предположении наличия всех цифр $x_i \in \overline{(0, k-1)}$) определяет самая невыгодная цифра, которая приводит к длинной цепочке обнуления счетчика, хранящего эту цифру, до нуля. Так, при $K=12$ и наборах шагов (1,2); (1,3); (1,4) и (1,6) этой цифрой будет 11; при наборе шагов (1,5)- цифра 9; при наборе шагов (1,5) - цифра 6. Число тактов преобразования для соответствующих наборов шагов при $K=12$ и $K=10$ приведено в табл.2.

Для 2-шагового ПК НКЭ с параллельной стратегией использования шагов преобразования число тактов преобразования можно найти в явном виде из формулы:

$$N_2^{\text{пар}}(K) = [(x_i - (a-1) \cdot \overline{x_i / a}], (a-1)], \quad x_i = \overline{(K-1), a}, \quad (2)$$

где K – основание системы счисления на входе; $[\]$ – округление до меньшего целого; a скобки $[\]$ – означают выбор большего из чисел, каждое из которых заключено в круглые скобки.

Таблица 1

Такт	К=12. Набор шагов (1,2)	К=12. Набор шагов (1,3)
	11 10 9 8 7 6 5 4 3 2 1 0	11 10 9 8 7 6 5 4 3 2 1 0
1	9 8 7 6 5 4 3 2 1 0 0 0	8 7 6 5 4 2 2 1 0 1 0 0
2	7 6 5 4 3 2 1 0 0 0 0 0	5 4 3 2 1 1 1 0 0 0 0 0
3	5 4 3 2 1 0 0 0 0 0 0 0	2 1 0 1 0 0 0 0 0 0 0 0
4	3 2 1 0 0 0 0 0 0 0 0 0	1 0 0 0 0 0 0 0 0 0 0 0
5	1 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0
6	0 0 0 0 0 0 0 0 0 0 0 0	
Такт	К=12. Набор шагов (1,4)	К=12. Набор шагов (1,5)
	11 10 9 8 7 6 5 4 3 2 1 0	11 10 9 8 7 6 5 4 3 2 1 0
1	7 6 5 4 3 2 1 0 2 1 0 0	6 5 4 3 2 1 0 3 2 1 0 0
2	3 2 1 0 2 1 0 0 1 0 0 0	1 0 3 2 1 0 0 2 1 0 0 0
3	2 1 0 0 1 0 0 0 0 0 0 0	0 0 2 1 0 0 0 1 0 0 0 0
4	1 0 0 0 0 0 0 0 0 0 0 0	0 0 1 0 0 0 0 0 0 0 0 0
5	0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0
Такт	К=12. Набор шагов (1,6)	К=12. Набор шагов (1,7)
	11 10 9 8 7 6 5 4 3 2 1 0	11 10 9 8 7 6 5 4 3 2 1 0
1	5 4 3 2 1 0 4 3 2 1 0 0	4 3 2 1 0 5 4 3 2 1 0 0
2	4 3 2 1 0 0 3 2 1 0 0 0	3 2 1 0 0 4 3 2 1 0 0 0
3	3 2 1 0 0 0 2 1 0 0 0 0	2 1 0 0 0 3 2 1 0 0 0 0
4	2 1 0 0 0 0 1 0 0 0 0 0	1 0 0 0 0 2 1 0 0 0 0 0
5	1 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 1 0 0 0 0 0 0
6	0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0

Таблица 2

Шаги	$N_2^{\text{пар}}(12)$	Шаги	$N_2^{\text{пар}}(10)$
1,2	6	1,2	5
1,3	5	1,3	4
1,4	5	1,4	4
1,5	5	1,5	5
1,6	6	1,6	5
1,7	6	1,7	6
1,8	7	1,8	7
1,9	8	1,9	8
1,10	9		
1,11	10		

3. Назначение и возможности программного средства "TRANSFORMATION"

К возможностям программного средства "TRANSFORMATION" относятся:

– имитация работы ПК НКЭ с числом шагов преобразования в наборе от 1 до 4, с основанием системы счисления на входе от 3 до 15 и разрядностью n преобразуемых цифр от 2 до 12;

– реализация двух режимов работы: режима преобразования заданного числа (или числа, сформированного генератором случайных чисел) и режима "СТАТИСТИКА" – для нахождения математического ожидания числа тактов преобразования и дисперсии.

Математическое ожидание и дисперсия определяются для всего множества преобразуемых чисел. Структурная схема алгоритма режима "СТАТИСТИКА" приведена на рис.1.

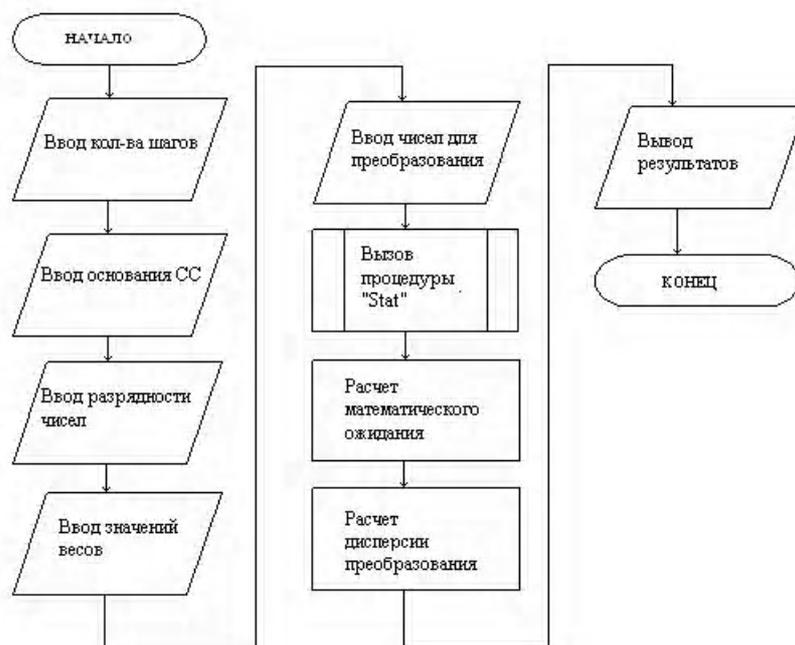


Рис. 1

Для разработки данного программного продукта была выбрана наиболее оптимальная по скорости проектирования и эффективности визуальная среда программирования QT для Windows XP/7, программа реализована на языке C++.

Для функционирования программы в составе компьютера необходимо иметь процессор с быстродействием не менее 1ГГц, 512 Мб ОЗУ, 50 м дискового пространства, Windows XP/Vista/7. Разработанное программное обеспечение сформировано как независимый модуль и для своего функционирования не требует дополнительного программного обеспечения. После запуска программного средства следует активизировать режим “ СТАТИСТИКА”. Код функции, реализующий режим статистики, приведен ниже:

```

do
{
int digit = 0;
for (int i = 0; i < amount_of_digits; ++i)
if (value[i] > digit)
digit = value[i];
int amountOfCycles = 0;
do
{
int equivalent = 0;
for (int j = 0; j < stepsCount; ++j)
{
quint8 step = stepsSizes.at(j).toInt();
if (step <= digit && step > equivalent)
equivalent = step;
}
digit -= equivalent;
if (equivalent == 0)
continue;
amountOfCycles++;
totalAmountOfCycles++;
}
while (digit > 0);
convertedValues[amountOfCycles]++;
}
  
```

```

amountOfCyclesPerConversation.append(amountOfCycles);
if (amountOfCycles > maxAmountOfCycles)
maxAmountOfCycles = amountOfCycles;
totalAmountOfValues++;
}

```

После активации режима “СТАТИСТИКА” появляется окно этого режима.

В данном окне имеется пять текстовых полей для ввода данных, три из которых снабжены кнопками инкремента и декремента. С помощью этих кнопок следует задать значение основания, количество преобразуемых цифр, число шагов, величину шагов.

Когда все значения введены, следует нажать поле “Собрать данные”. Программа после вычисления отобразит значения математического ожидания, дисперсии, количества всех проанализированных чисел и распределение количеств чисел по числу тактов преобразования, на основании которого можно построить гистограмму распределения.

Результат работы программного средства в режиме “СТАТИСТИКА” для $K=10$ и $n=6$ приведен на рис. 2.

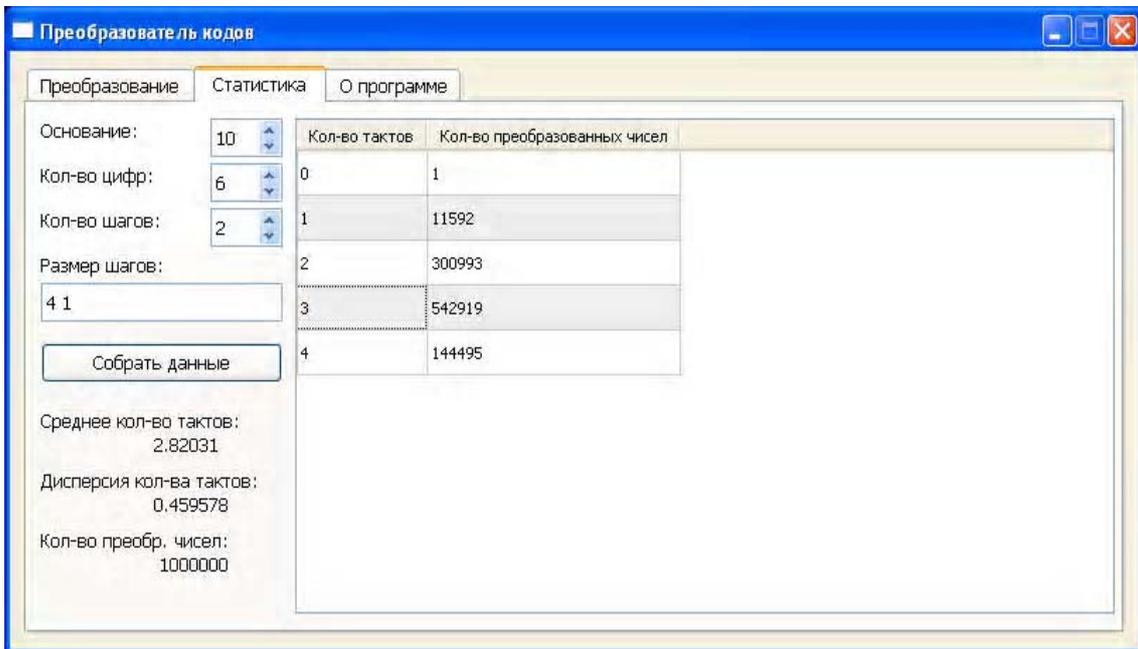


Рис. 2

Для сравнения результатов статистики при тех же значениях $K=10$; $n=6$ для последовательной стратегии использования шагов с помощью программного средства “CONVERTOR” получены аналогичные статистические данные. Результаты статистики для обеих стратегий в целях сравнения сведены в табл.3.

Таблица 3

Число тактов	Тип стратегии	
	Последовательная	Параллельная
0	1	1
1	126	11 592
2	5 299	300 993
3	87 157	542 919
4	385 049	144 495
5	522 368	

В основной части табл.3. приведено распределение количеств чисел по числу тактов преобразования. Одно число (000 000) преобразуется за 0 тактов; общее количество чисел составляет $N = 10^6$. Статистические данные для M посл ; D посл ; M пар ; Nпар ; a

также значения коэффициента $Y = M_{\text{пар}} / M_{\text{посл}}$ для ряда оснований $K=3-12$ приведены в табл. 4.

Имеется возможность задавать различное число шагов преобразования - от 1 до 8, основание системы счисления на входе преобразователя - от 3 до 15, разрядность чисел - от 2 до 12, веса шагов преобразователя. Кроме этого, данный режим позволяет вручную задавать набор преобразуемых чисел или автоматически на выбор. При автоматическом формировании чисел программа делает полный перебор всех возможных значений в диапазоне от 0 до $n-1$, где n - разрядность преобразуемых чисел.

Таблица 4

К	3	5	6	7	9	10	11	12
a	2	4	4	4	4	4	4	4
$M_{\text{посл}}$	1,82	2,19	3,49	3,53	4,25	4,42	4,52	4,72
$a_{\text{посл}}$	0,15	0,23	0,46	0,33	0,46	0,25	0,37	0,27
$M_{\text{паралл}}$	0,998	1,78	1,823	2,534	2,643	2,82	3,334	4,012
$a_{\text{паралл}}$	0,146	1,725	0,289	0,446	0,811	0,46	0,601	1,026
γ	0,548	0,729	0,521	0,882	0,626	0,638	0,737	0,85

Математическое ожидание M рассчитывается по формуле:

$$M = \left(\sum_{i=0}^{i=N_{\max}} i \cdot N_i \right) / 2^n, \quad (3)$$

где M - математическое ожидание числа тактов преобразования i ; N_i - количество чисел, преобразуемых за i тактов; N_{\max} - максимальное значение числа тактов преобразования (зависит от основания системы счисления K и значений шагов преобразования a, b, c); n - число входных преобразуемых разрядов.

Затем рассчитывается дисперсия преобразования D по формуле:

$$D = \left(\sum_{j=1}^{j=2^n} (M - N_j)^2 \right) / 2^n (2^n - 1), \quad (4)$$

где N_j - количество тактов, за которое преобразуется число $A(K)$.

Анализ табл. 4 показывает, что параллельная стратегия использования шагов дает существенный выигрыш по быстродействию в 35 - 45% для $K=3-10$ и 15% для $K=12$.

Выводы

Научная новизна выполненного исследования заключается в следующем.

1. Рассмотрена и проанализирована параллельная стратегия использования шагов преобразования в многошаговых ПК НКЭ

2. Получена формула для вычисления числа тактов преобразования 2-шагового ПК НКЭ параллельного типа.

3. Проанализировано назначение и возможности программного средства "TRANSFORMATION" в режиме "СТАТИСТИКА".

4. Проведен сравнительный анализ статистических характеристик (математического ожидания и дисперсии) в ПК НКЭ для двух стратегий - последовательной и параллельной.

Практическая значимость исследования состоит в доказательстве преимущества параллельной стратегии перед последовательной и в возможности быстрого автоматизированного расчета основных статистических характеристик ПК НКЭ с параллельной стратегией.

Список литературы: 1. А.С. 1647908 5НОЗМ 7/12. Преобразователь двоично- K -ичного кода в двоичный код / Н.Я.Какурин, Ю.К.Кирьяков, А.Н.Макаренко // Открытия, изобретения. 1991. №17. С. 262-263. 2. А.С. 1783618 5НОЗМ 7/12. Преобразователь двоично- K -ичного кода в двоичный код / Н.Я.Какурин, А.Н.Макаренко, Д.Ю.Исхаков, В.А.Толмацкий // Открытия, изобретения. 1984. №44. С. 250. 3. Какурин Н.Я., Лопузин Ю.В., Быкова Н.Н. Программное средство для анализа преобразований чисел // АСУ и приборы автоматики. 2007. Вып.140. С.96-102. 4. Какурин Н.Я., Вареца В.В., Коваленко С.Н. Параллельная стратегия использования шагов в двухшаговых преобразователях кодов // АСУ и приборы автоматики. 2007. Вып.141. С.29-36.

Поступила в редколлегию 23.05.2011

Лопухин Юрий Владимирович, ст. преподаватель кафедры АПВТ ХНУРЭ. Научные интересы: проектирование программного обеспечения, автоматизация проектирования цифровых устройств. Адрес: Украина, 61166, Харьков, пр.Ленина, 14, тел. 70-21-326.

Бочаров Евгений Витальевич, студент группы КИ-07-6 ХНУРЭ. Научные интересы: автоматизация проектирования цифровых устройств, проектирование программного обеспечения. Адрес: Украина, 61166, Харьков, пр.Ленина, 14, тел. 70-21-326.

Вареца Виталий Викторович, аспирант кафедры АПВТ ХНУРЭ. Научные интересы: проектирование программного обеспечения, автоматизация проектирования цифровых устройств. Адрес: Украина, 61166, Харьков, пр.Ленина, 14, тел. 70-21-326.

Полежаев Кирилл Вадимович, студент группы СИ-07- ХНУРЭ. Научные интересы: автоматизация проектирования цифровых устройств, проектирование программного обеспечения. Адрес: Украина, 61166, Харьков, пр. Ленина, 14, тел. 70-21-326.

Макаренко Анна Николаевна, канд. техн. наук, доцент кафедры информационных технологий Харьковского института банковского дела. Научные интересы: информационные технологии, анализ и синтез преобразователей код-код. Адрес: Украина, 61074, Харьков, пр. Победы, 55, тел. 336-05-64.

ПОСТРОЕНИЕ АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ УПРАВЛЕНИЯ ПРОИЗВОДСТВЕННЫМИ ЗАКАЗАМИ НА ОСНОВЕ АРХИТЕКТУРЫ ОПЕРАЦИОННЫХ СИСТЕМ. ЧАСТЬ 2

Экспертная система управления заказами (ЭСУЗ) в логистической сети предприятия как система программной поддержки гибкого производства реализует архитектуру операционных систем. Описываются функции, требования к управлению и построению системы. Управление заказами включает одноразовые операции (состав заказов, набор операций, блок управления заказами, разбивка заказов по участкам/службам предприятия) и много-разовые операции (запуск, приостановка, блокирование, разблокирование заказов), а также планирование и взаимодействие заказов. Во второй части рассматриваются требования к экспертной системе управления заказами и ее построение.

1. Как управлять заказами. Требования к ЭСУЗ*

Диаграмма состояний заказов приведена на рисунке.



* Требования к программе, управляющей заказами в логистической сети предприятия, сформулированы на основе архитектуры операционных систем.

На рисунке сноски означают:

¹ Отсутствуют: материалы (истекли, пришли в негодность, использованы для экстренных заказов); оборудование (отказало); необходимый персонал. Ожидается аннулирование этого заказа; необходима переналадка оборудования и (или) оснастки или нужна консультация с заказчиком.

² 1) Истекло время, отведенное для выполнения заказа. В этом случае ЭСУЗ идентифицирует заказ и в диалоговом режиме предлагает начальнику участка ввести комментарии (причины, пути их преодоления), а диспетчеру – принять решение о дальнейшем выполнении заказа (возможность, целесообразность, прекращение выполнения и действия предприятия в этом случае. 2) Изменились приоритеты заказов в очереди. 3) Поступило извещение об изменении заказа, например, об изменении ассортимента заказанной продукции.

Всякий новый появляющийся заказ должен быть представлен диспетчеру в форме договора на поставку или спецификации с указанием необходимых сведений. Диспетчер размещает заказ в ЭСУЗ (состояние «рождение»). Новый заказ автоматически вводится в финансово-учетный контур предприятия. При рождении осуществляется обработка заказа. Он получает ресурсы, устанавливается значение программного счетчика этого заказа. Программный счетчик содержит информацию о том, какие этапы производства (техпроцессы) должны быть выполнены, или число операций какого-то техпроцесса, или сколько штук сделано (для партий деталей), или какие этапы проектирования (разработки конструкторской документации) должны быть выполнены (с указанием событий – вех).

Обработка опытных заказов включает также: 1) согласование конструкторской документации с начальником (или технологом) вовлеченных производственных участков на предмет возможности изготовления, технологичности и идентификацию требований, когда они не четко специфицированы документально; 2) согласование времени выполнения заказа и другие сервисные процедуры (время и условия доставки готовой продукции заказчику).

Родившийся заказ переводится в состояние «готовность». На этом этапе осуществляется планирование выполнения заказов, запуск в производство и блокирование (в случае необходимости). ЭСУЗ, пользуясь каким-либо алгоритмом планирования, устанавливает очередь из готовых заказов и переводит их в состояние «исполнение». Покинуть состояние «исполнение» заказ может по трем причинам:

- он заканчивает свою деятельность (выполнен);
- он не может продолжать свою работу (блокирование), пока не произойдет некоторое событие и ЭСУЗ переводит его в состояние «ожидание»;
- в результате возникновения прерывания (прерывание диспетчером или самой ЭСУЗ: истекло время, др.) его возвращают в состояние «готовность».

В случае если истекло время выполнения заказа, ЭСУЗ предлагает диспетчеру принять решение и перевести его в одно из возможных состояний: 1) «исполнение» (при этом заказ идентифицируется как задержанный); 2) «ожидание» (для консультаций с заказчиком); 3) «закончил исполнение». В состоянии «ожидание» ЭСУЗ периодически посылает системные вызовы (запросы) диспетчеру с предложением изменить состояние заказа. При завершении своей деятельности заказ из состояния «исполнение» попадает в состояние «закончил исполнение».

Заказ не может сам перейти из одного состояния в другое. Изменением состояния заказов занимается ЭСУЗ, совершая операции над ними в диалоговом режиме. Количество таких операций совпадает с количеством стрелок на диаграмме состояний. Можно выделить три парные операции:

- создание заказа – завершение заказа;
- приостановка заказа (перевод из состояния «исполнение» в состояние «готовность») – запуск заказа (перевод из состояния «готовность» в состояние «исполнение»);
- блокирование заказа (перевод из состояния «готовность» или «исполнение» в состояние «ожидание») – разблокирование заказа (перевод из состояния «ожидание» в состояние «готовность»).

Существует еще одна операция, не имеющая парной: изменение приоритета заказа. Операции создания и завершения заказа являются одноразовыми, так как применяются к

нему не более одного раза. Все остальные операции, связанные с изменением состояния заказов, как правило, являются многоазовыми.

Каждый заказ представляется в ЭСУЗ в виде некоторой структуры данных. Эта структура – блок управления заказом (БУЗ) – содержит информацию, специфическую для данного заказа:

- состояние, в котором находится заказ (готовность, ожидание, исполнение);
- программный счетчик заказа или, другими словами, ежедневные или еженедельные задания, которые должны быть выполнены;
- данные по требуемым ресурсам;
- данные, необходимые для планирования: приоритет заказа, его размер, потребитель (получатель);
- учетные данные: идентификационный номер заказа, какой пользователь инициировал его работу, общее время использования ЭСУЗ данным заказом (счетчик);
- информацию об оборудовании, связанном с заказом: таблица работоспособного (в наличии) оборудования – заполняется начальниками участков в диалоговом режиме;
- комментарии по прерываниям.

БУЗ создается для каждого заказа. Начальники производственных участков получают доступ для редактирования той части БУЗ, которая содержит данные соответствующего участка. Прочие – только для чтения. Итак, вся информация, необходимая для совершения операций над заказами, доступна ЭСУЗ. Она хранится в БУЗ, который является моделью заказа для ЭСУЗ. Любая операция, производимая в ЭСУЗ над заказом, вызывает определенные изменения в БУЗ.

Заказы для участков создаются динамически по мере поступления внешних заказов. Инициатором разделения заказов может быть диспетчер либо – сама ЭСУЗ, т.е. некоторая процедура в ЭСУЗ, описанная алгоритмически с привлечением экспертов – начальников участков, технологов, конструкторов, др.

Внешний заказ, инициировавший новый (внутренний) будем называть заказом-родителем, а вновь созданный заказ будем называть заказом-ребенком. Заказы-дети могут, в свою очередь, порождать новых детей, т.е. заказы для служб и отделов предприятия. Это: заказы на маршрутные листы и другую технологическую документацию; на разработку и изготовление оснастки и (или) нестандартного оборудования; на материалы и покупные изделия или на склады об отпуске материалов; заказы в инструментальный склад; на услуги отдела технического контроля (планирование объема проверок и контроля).

При рождении нового заказа диспетчер заводит новый БУЗ с состоянием «рождение» и начинает его заполнение. Новый заказ получает свой собственный уникальный идентификационный номер. Для выполнения заказа необходимы ресурсы. Существуют два подхода к их выделению. Новый заказ может получить в свое распоряжение некоторую часть ресурсов, запланированных и приобретенных на основании перспективных планов-заказов от заказчиков. При этом заказ разделяет права на ресурсы с другими заказами или может получить ресурсы в виде материалов, не израсходованных ранее или приобретенных специально для текущего заказа. Информация о выделенных ресурсах заносится в БУЗ самой ЭСУЗ по расчетной программе для типовых деталей и подтверждается начальниками участков в режиме диалога.

После того, как заказ наделен содержанием и в БУЗ записывается необходимая информация, состояние нового заказа меняется на «готовность». После того, как заказ завершен, ЭСУЗ переводит его в состояние «закончил исполнение» и делает соответствующую запись в БУЗ. При этом сам БУЗ не уничтожается, а остается в ЭСУЗ еще некоторое время. Это связано с тем, что диспетчер после завершения заказа может запросить ЭСУЗ о причине завершения заказа и (или) статистическую информацию о его выполнении. В отношении уже произведенной продукции ЭСУЗ запрашивает диспетчера и помещает данные о ней либо в задел, либо в блок «утилизация» (использования в качестве вторичного сырья), или в блок «уничтожение».

Выше были описаны одноразовые операции, которые осуществляет ЭСУЗ – такие операции, которые приводят к изменениям количества заказов, находящихся под управлением ЭСУЗ. Они всегда связаны с выделением и освобождением определенных ресурсов.

ЭСУЗ может осуществлять многоразовые операции – они не приводят к изменению количества заказов и не обязаны быть связанными с выделением или освобождением ресурсов. Действия ЭСУЗ при выполнении многоразовых операций над заказами следующие.

Запуск заказа. Из числа заказов, находящихся в состоянии «готовность», ЭСУЗ формирует (по приоритетам) заказы для последующего исполнения. Критерии и алгоритмы такого выбора будут рассмотрены ниже. Для выбранного заказа ЭСУЗ выделяет ресурсы, необходимые для его выполнения (удаляет часть ресурсов из каталогов). БУЗы, после их заполнения диспетчером, отправляются начальникам участков для согласования, а также обеспечивающим службам для планирования своей работы. После утверждения БУЗов начальниками участков состояние заказа меняется на «исполнение».

Приостановка заказа. Работа по заказу, находящемуся в состоянии «исполнение», приостанавливается в результате какого-либо прерывания. Начальник вовлеченного участка по запросу ЭСУЗ сохраняет программный счетчик – вводит необходимую информацию по специальному адресу обработки данного прерывания. По указанному адресу обычно располагается одна из частей ЭСУЗ. Она сохраняет динамическую часть заказа в его БУЗе, переводит заказ в состояние «готовность» и диспетчер приступает к выполнению определенных действий, связанных с возникшим прерыванием.

Блокирование заказа. Заказ блокируется, когда он не может начать или продолжить свою работу, не дождавшись возникновения какого-либо события. ЭСУЗ добавляет заказ в очередь заказов, ожидающих освобождения людей, оборудования или возникновения других событий, и, при необходимости, сохранив нужную часть структуры данных заказа в его БУЗе, переводит заказ из состояния «готовность» или «исполнение» в состояние «ожидание».

Разблокирование заказа. После возникновения какого-либо события ЭСУЗ определяет, в диалоговом режиме с диспетчером, какое именно событие произошло. Затем проверяет: находился ли некоторый заказ в состоянии «ожидание» для данного события и, если находился, переводит его в состояние «готовность». При этом диспетчер выполняет необходимые действия, связанные с наступлением события (инициализация операции запуска в работу, изменение приоритета).

Деятельность ЭСУЗ состоит из цепочек операций, выполняемых над различными заказами, и сопровождается переключением работы производственных участков и служб с одного заказа на другой. Например, при выполнении участком некоторого заказа возникает прерывание от диспетчера или самой ЭСУЗ. Над выполняющимся заказом производится операция приостановки. Далее диспетчер (в диалоговом режиме с ЭСУЗ) осуществляет запуск приостановленного или нового заказа, выбранного в соответствии с приоритетом планирования. Таким образом, в результате обработки информации о прерывании возможна замена в заказах, находящихся в состоянии «исполнение» (очередность заказов). Для корректного переключения производственного участка с одного заказа на другой необходимо сохранить контекст исполнявшегося заказа и восстановить контекст заказа, на который будет переключен участок. Восстановление контекста заказа предусматривает переналадку оборудования и оснастки. Время, затраченное на это, не используется участком для совершения полезной работы и представляет собой накладные расходы, снижающие производительность.

Существуют два уровня планирования (заказов-родителей): долгосрочное и краткосрочное (планирование выполнения).

Долгосрочное планирование осуществляется следующим образом: ЭСУЗ (в соответствии со специальным программным кодом) посылает запросы потенциальным предприятиям-заказчикам о планируемых работах. Запросы посылаются в согласованные сроки и в согласованной форме. При этом желательно равномерное распределение заказов-родителей. ЭСУЗ призвана, по возможности, влиять на интенсивность размещения заказов клиентами, чтобы предотвратить «комки заказов», почти одновременное поступление заказов от многих заказчиков [1]. Это осуществляется путем системных вызовов потенциальным заказчикам с предложением сделать заказ. При этом ЭСУЗ может руководствоваться утвержденными планами и инновационными проектами. Ответы на посылаемые системные вызовы должны использоваться для корректировки в диалоговом режиме календарного плана производства.

Планирование выполнения заказов выступает в качестве краткосрочного планирования. Оно проводится, к примеру, при возникновении прерываний или при поступлении очередного заказа. Выбор новой очередности выполнения заказов влияет на функционирование ЭСУЗ до поступления очередного заказа. В некоторых случаях бывает необходимо, или выгодно, изменить приоритет частично выполненного заказа, т.е. остановить (прервать), чтобы завершить позже. Когда и какой из заказов нужно остановить и запустить снова, решается дополнительным промежуточным уровнем планирования заказов – среднесрочным.

Планирование выполнения заказов включает критерии планирования и правила. Выбор правил планирования определяется классом задач и целями, которых мы хотим достичь, используя планирование. К числу целей можно отнести, например, сокращение полного времени выполнения или времени ожидания.

Независимо от поставленной цели планирования желательно, чтобы правила обладали следующими свойствами:

- Были предсказуемыми. Для этого одни и те же задания рабочим должны выполняться приблизительно за одно и то же время.

- Минимальные накладные расходы, связанные с планированием. Если на каждые 100 часов, выделенных заказу, будет приходиться 200 часов на определение того, какой именно заказ будет приоритетным и на переключение контекста (переналадка оборудования), то такие правила, очевидно, использовать не стоит.

- Равномерная загрузка ресурсов производственных участков (предпочтение отдается заказам, которые будут занимать малоиспользуемые ресурсы).

Правила планирования должны опираться на какие-либо характеристики заказов, состояния самих производственных участков и служб, иными словами, на параметры планирования. Все параметры планирования можно разбить на две большие группы: статические и динамические. Статические параметры не изменяются в ходе функционирования производства, динамические же, напротив, подвержены постоянным изменениям.

К статическим параметрам производственных участков можно отнести предельные значения их ресурсов (оборудование, оснастка, персонал, площади, др.).

К статическим параметрам заказов относятся характеристики, как правило, присущие уже на этапе запуска:

- Каким заказчиком запущен заказ.
- Насколько важным является поставленный заказ (каков приоритет его выполнения).
- Сколько времени запрошено заказчиком для выполнения заказа.
- Каково соотношение производственного времени и времени, необходимого для осуществления подготовки производства.

- Какие ресурсы производственного участка и в каком количестве необходимы заказу.

Для среднесрочного планирования в качестве динамических параметров заказов может выступать следующая информация:

- Сколько времени прошло с момента запуска заказа или его перевода в состояние «готовность».

- Какую часть общего месячного плана составляет заказ.

- Сколько ресурсов (материалов, станков, персонала и др.) уже предоставлено заказу.

Динамические параметры производственных участков и служб описывают количество свободных ресурсов в текущий момент времени. Для краткосрочного планирования понадобится ввести еще два динамических параметра. Деятельность любого заказа можно представить как последовательность циклов использования оборудования и ожидания завершения операций переналадки оборудования и оснастки: соответственно промежуток времени непрерывного использования оборудования (НИО) и промежуток времени непрерывного ожидания переналадки (НОП). Значения продолжительности последних и очередных НИО и НОП являются важными динамическими параметрами заказа. Они должны измеряться и обновляться.

Правила краткосрочного планирования, которые должен обеспечивать МПВЗ:

- 1) Обычное правило планирования – первым пришел, первым выполнен (ПППВ).
- 2) Кратчайшая работа первой (КРП).
- 3) Приоритетное планирование.

Порядок расположения заказов в очереди является важным. Если небольшие заказы расположены в очереди ближе к ее началу, то общая производительность возрастает. Если бы мы знали время следующих НИО для заказов, находящихся в состоянии «готовность», то могли бы выбрать для выполнения не заказ из начала очереди, а заказ с минимальной длительностью НИО. Если же таких заказов два или больше, то для выбора одного из них можно использовать алгоритм ПППВ.

КРП-правило краткосрочного планирования может быть как вытесняющим, так и невытесняющим. При невытесняющем КРП-планировании оборудование предоставляется заказу на все требующееся ему время, независимо от появления новых заказов. При вытесняющем КРП-планировании учитывается появление новых заказов в очереди (из числа вновь родившихся или разблокированных) во время выполнения выбранного заказа. Если НИО нового заказа меньше, чем его остаток у исполняющегося, то исполняющийся заказ вытесняется новым. Основную сложность при реализации правила КРП представляет невозможность точного знания времени очередного НИО для исполняющихся заказов. При краткосрочном планировании мы можем делать только прогноз длительности следующего НИО, исходя из предыстории выполнения заказов (это может обеспечить ЭСУЗ, что является одним из ее преимуществ).

Правило КРП – это частный случай приоритетного планирования. При приоритетном планировании каждому заказу присваивается определенное числовое значение – приоритет, в соответствии с которым он выполняется. Заказы с одинаковыми приоритетами планируются в порядке ПППВ. Для правила КРП в качестве такого приоритета выступает оценка продолжительности следующего НИО. Чем меньше значение этой оценки, тем более высокий приоритет имеет заказ. В общем случае принципы назначения приоритетов могут опираться как на внутренние критерии производства, так и на внешние по отношению к нему. Внутренние используют различные количественные и качественные характеристики заказа для вычисления его приоритета. Это могут быть, например, ограничения по времени использования оборудования; требования к размеру склада; число задействованных единиц оснастки, станков; отношение средних продолжительностей НОП и НИО. Внешние критерии исходят из таких параметров, как важность заказа для достижения каких-либо целей, стоимость оплаченной работы и других факторов.

Планирование с использованием приоритетов также может быть как вытесняющим, так и невытесняющим. При вытесняющем планировании заказ с более высоким приоритетом, появившийся в очереди заказов, вытесняет выполняемый заказ с более низким приоритетом. В случае невытесняющего планирования он просто становится в начало очереди заказов.

Существуют причины для взаимодействия заказов во время их параллельного выполнения.

Одной из причин является повышение скорости работы. Когда один заказ ожидает наступления некоторого события, другие в это время могут выполняться.

Другой причиной является совместное использование технологического оборудования. К примеру, изделия различных заказов могут одновременно обрабатываться на участке гальваностегии.

Также это необходимо для удобства работы диспетчера, желающего редактировать и отлаживать заказы одновременно.

Связь может быть установлена путем автоматической адресации данных о заказах в некоторый промежуточный объект со своим интерфейсом (файл заказов) для хранения данных.

Файл заказов должен содержать сигнальную информацию об исполнении этапов, соответствующих программным счетчикам всех заказов. Этапы (осуществление одного из предусмотренных производственных процессов) должны относиться к какому-либо документу: протокол, чертеж, технологическая карта, маршрутный лист, накладная, счет.

Составные части файла заказов:

- таблица видов этапов;
- таблица, в которой указано какой документ содержит этот этап;
- таблица, в которую заносят исполнение этапов при прохождении заказа.

Все таблицы должны быть соответствующим образом связаны.

2. Построение ЭСУЗ

В базе знаний в некотором закодированном виде хранятся формализованные знания экспертов – специалистов в соответствующей предметной области. Существуют различные формы представления знаний в экспертных системах. Наиболее простая – это форма представления, ориентированная на объекты [2]. Эта форма определяет: (1) *объект*, обладающий (2) *свойствами*, которые могут принимать (3) *значения* из известного набора. Эта форма представления знаний является наиболее ранней, своего рода переходной, от представления данных к представлению знаний. Однако представление, ориентированное на объекты, является достаточным для решения задач планирования и наблюдения заказов.

Форма представления знаний в ЭСУЗ приведена в табл. 1.

Таблица 1

Заказы*	Состояние заказа	Работы (операции)
1	2	3
ИН (n)	n.1. Рождение	Обработка заказа
		n.1.1. Конструкторская документация согласована и принята начальником участка (технологом)
		n.1.2. Время выполнения заказа согласовано
		n.1.3. Время и условия доставки заказчику готовой продукции согласованы
		n.1.4. Спецификация согласована и принята диспетчером
		n.1.5. БУЗ заполнен**
	
		Передача заказа
		n.1.6. Сообщение об отклонении заказа или о возражении по условиям заказчика не отправлено
		n.2. Готовность
	n.2.1. Очередь выполнения заказов установлена	
	
	Запуск в производство	
	n.2.2. Заказы – дети разосланы для производства работ	
	n.2.3. Контекст ранее прерванного заказа восстановлен	
	
	Блокирование заказа	
	n.2.4. Заказ заблокирован; причина блокирования, а также событие, необходимое для разблокирования указаны	
	n.3. Исполнение	
		n.3.1. Исполнение заказа приостановлено; причина прерывания указана
		n.3.2. Программный счетчик сохранен
		n.3.3. Комментарии по прерыванию введены в БУЗ
		n.3.4. Решение о дальнейшем выполнении заказа принято
		n.3.5. Контекст прерванного заказа сохранен
	
		Блокирование
		n.3.6. Заказ заблокирован; причина блокирования, а также событие, необходимое для разблокирования указаны
n.3.7. Необходимая часть структуры данных заказа сохранена в БУЗ		
....		
		Операции технологических процессов
		n.3.8. Заготовительная
		n.3.9. Токарная обработка

1	2	3
		n.3.10. Фрезерная обработка
		n.3.11. Слесарная обработка
	
		n.3.12. Подготовка производства
		n.3.13. Получение и хранение материалов
		n.3.14. Получение и хранение инструмента
	
		Завершение работы
	n.3.15. Все работы по заказу выполнены	
	n.4. Ожидание	n.4.1. Заказ разблокирован (ожидаемое событие произошло)
	
	n.5. Закончил исполнение	Завершение исполнения
		n.5.1. Готовая продукция доставлена заказчику

* Должны быть указаны уникальные идентификационные номера (ИН) заказов: $n=1,2, \dots, N$, т.е. всего в исполнении N заказов.

** В том числе ресурсы выделены, их наличие согласовано с производством, со складом; значения программного счетчика заданы; таблица работоспособного оборудования актуализирована.

Здесь объекты – это заказы; свойства – это состояния заказов; допустимые значения свойств – это работы, множество работ (или операций) производственных процессов (технологических, контрольных, отгрузочных, складских, транспортных), которые выполняются в рамках соответствующего состояния заказа. Механизм планирования выполнения заказов использует форму обработки событий, приведенную в табл. 2.

Таблица 2

События	Действия	Новые состояния заказов*	Приоритет
Поступление нового заказа с номером ИН (n)	Согласовать конструкторскую документацию	ИН (n) = n.1.1.	P ₁
	Согласовать сервисные атрибуты заказа	ИН (n) = n.1.2.	P ₂
		ИН (n) = n.1.3.	P ₃
	Согласовать спецификацию	ИН (n) = n.1.4.	P ₄
	Ввести необходимые данные в БУЗ	ИН (n) = n.1.5.	P ₅
	Принять заказ	ИН (n) = n.1.6.	P ₆
	Корректировать приоритеты выполнения готовых заказов**.	ИН (n) = n.2.1.	P ₇
		Использовать невытесняющее КРП-планирование	ИН (n-1) = (n-1).2.1.
	Заказы – дети приоритетного заказа – родителя разосланы производственным участкам	ИН (n) = n.2.2.	P ₉

* Указывают выполненные работы (операции).

** Здесь всего в состоянии «готовность» два заказа: с номерами ИН (n), ИН (n-1); первым выполняют заказ ИН (n).

Важным этапом создания ЭСУЗ является заполнение экспертами табл. 1 (столбец «работы») и табл. 2 (приведен только фрагмент), а также мониторинг за происходящими событиями – он осуществляется вовлеченным персоналом как ответы на системные вызовы ЭСУЗ. Особенно подробно должна быть составлена табл. 2, так как ЭСУЗ будет реагировать на все события посредством действий, предусмотренных этой таблицей. Табл. 2 должна быть открыта для корректировки и добавлений, возможных исходя из опыта работы. Она содержит приоритеты для всех событий и связанных с ними действий, поскольку одновременно могут произойти два события или более, или происходит одно событие, в то время как ЭСУЗ обрабатывает другое.

Заключение

Стратегия гибкого производства предусматривает размещение потребителя в центре процесса планирования деятельности предприятия. Учет состояния внешней среды (событий, возникающих в производстве) предопределяет более тесное (чем обычно, даже в условиях стандартов MRP II и CSRP) взаимодействие потребителя с бизнес-процессами предприятия. Речь идет о преодолении неожиданных обстоятельств (показатель бесперебойности логистического цикла), особых действиях по исправлению срывов и недостатков (показатель уровень брака / устранение недостатков), оптимизации некоторых работ универсального производства, т.е. необходимо выбрать альтернативу с учетом состояния среды. При этом управление в значительной мере основывается на опыте, интуиции и здравом смысле.

Такую деятельность невозможно описать и автоматизировать без привлечения интеллектуальных программных средств, способных обучаться на основании входной информации и опыта. Формализация интеллекта принципиально ограничена. Однако некоторые интеллектуальные функции, например, эвристические (творческий опыт), могут быть осуществлены в экспертных системах. Экспертная система управления заказами призвана реализовать и помочь расширить творческие способности человека, а также передать программе определенные формализуемые умственные операции, выполняемые человеком. Указанная целевая направленность построения ЭСУЗ не отделена от другого важного аспекта – накопления знаний. Эта система, понятная для всех ключевых сотрудников организации, предназначена для обобщения и хранения знаний ее лучших работников.

Список литературы: 1. Джонсон Д., Вуд Д., Вордлоу Д., Мерфи - мл. П. Современная логистика / Джеймс С. Джонсон, Дональд Ф. Вуд, Дэниел Л. Вордлоу, Поль Р. Мэрфи-мл.; пер. с англ. М.: Издательский дом «Вильямс», 2004. 624 с. 2. Уотермен Д. Руководство по экспертным системам / Д. Уотермен; пер. с англ. М.: Мир, 1989. 388 с.

Поступила в редколлегию 28.02.2011

Ковалев Алексей Иванович, канд. техн. наук, начальник отдела управления проектами ОАО ЭК «Хмельницкоблэнерго». Научные интересы: управление предприятиями, оценивание деятельности, международные стандарты на системы управления, качество управления, информационные системы. Адрес: Украина, 29016, Хмельницкий, ул. Храновского, 11а, тел. (0382) 78-78-23.

УДК 004.031

В.М. ЛЕВИКІН, О.П. КОСТЕНКО, Є.Г. ЗІНЧЕНКО

РОЗРОБКА МОДЕЛІ КОМПЛЕКСУ ПРОЕКТНИХ ПАТЕРНІВ МАРКЕТИНГОВИХ ІНФОРМАЦІЙНИХ СИСТЕМ

Проводиться аналіз моделей та умов формування вимог до маркетингових інформаційних систем. Пропонується комплекс проектних патернів у вигляді набору різних модулів, що дозволяють розроблювати еталонну архітектуру системи з її представленням на різних рівнях. Описується модель компонентної діаграми проектних шаблонів розподіленої маркетингової інформаційної системи. Наводиться процедура реалізації модуля синхронізації маркетингових даних, які знаходяться у різних модулях системи.

1. Вступ

Для розробки багаторівневої розподіленої маркетингової інформаційної системи (МІС) застосовують топологічні моделі для її адекватних архітектурних рішень, які відображають значущу роль функціональних вимог, котрі пред'являються до системи. Існують загальні функціональні топологічні моделі вимоги, які повинні бути враховані при розробці подібних систем, а з урахуванням проблемно-орієнтованого характеру даної багаторівневої розподіленої МІС вони можуть бути додатково уточнені таким чином [1]:

– у системі повинен бути забезпечений постійний і оперативний доступ до необхідної інформації про всі основні маркетингові процеси (МП), ресурси та маркетингові (конкурентні) об'єкти у ринковому регіоні, представленої у наочній візуальній формі;

– цей доступ (у санкціонованому режимі) повинен бути можливий з будь-якої точки території ринкового регіону, як із стаціонарних комп'ютерів, так і, у перспективі, з переносних терміналів (ноутбуків, мобільних телефонів і т.п.);

- одним з сервісів системи повинна бути можливість постійно підтримувати оперативний зв'язок (шляхом передачі повідомлень електронною поштою) між різними категоріями її корпоративних користувачів: маркетинговим персоналом окремих центрів продажу маркетингових об'єктів, маркетологами регіональних (ринкових) сервісних центрів, керівництвом у центральному офісі маркетингового рівня підприємства і т.п.;

- система дозволить накопичувати, систематизувати і аналізувати будь-яку інформацію, необхідну для підготовки і прийняття оперативних і стратегічних рішень з контролю і управління маркетинговими ресурсами підприємств, а також для формування усіх видів сервісних центрів і центрів продажу.

Мета роботи – розробити моделі комплексу проектних патернів МІС.

2. Результати дослідження

При розробці архітектури МІС доцільно застосувати підхід, заснований на методі створення колекції шаблонів проектних рішень (ПР), а потім, на їх основі, і так званих повторно-використовуваних програмних компонентних рішень (ПВПКР), що забезпечує універсальний характер процесу проектування і реалізації усієї системи у цілому. Враховуючи функціональні вимоги, котрі пред'являються до маркетингової системи і мають певну специфіку збору і обробки маркетингових даних у МІС [2, 3], пропонуємо проектні шаблони у вигляді таких типових ПР (і відповідні програмні компоненти) [4]:

1. Інтегрована База Маркетингових Даних Модуля (ІБМДМ): цей компонент призначений для зберігання різної інформації, необхідної для забезпечення всієї функціональності відповідного WEB-ПК.

2. Сервіс Обміну Маркетинговими Даними (СОМД): компонент, який повинен підтримувати коректний обмін даними з модулями МІС, з урахуванням наявності декількох клієнтських додатків, які працюють у кожному модулі системи.

3. Сервіс Візуалізації Маркетингових Даних (СВМД): підмножина компонентів (включаючи серверні і клієнтські компоненти), що забезпечують для фахівців-маркетологів візуалізацію параметрів МП, різних системних повідомлень і тому подібне.

4. Сервіс Синхронізації Маркетингових Даних (ССМД): підмножина компонентів (серверні і клієнтські), які дозволяють регулярно проводити синхронізацію маркетингових даних, котрі зберігаються у різних модулях системи.

5. Менеджер Облікових Записів Фахівців-Маркетологів (МОЗФМ): компонент, за допомогою якого у системі відбувається реєстрація її користувачів, створюються їх групи, визначаються права доступу до даних для кожної групи.

Ці компоненти комплексу проектних патернів, які є складовими еталонної архітектури (ЕА) МІС, можуть бути представлені у нотації UML у вигляді компонентної діаграми, яка показана на рис. 1.

На ЕА компоненти згруповані у UML-пакети [5], які агрегують у собі функціональність трьох найбільш можливих типів модулів системи:

1) Модуль продуктового рівня, тобто окремого продажу промислової продукції; його характерною особливістю є наявність підсистеми збуту товарів та підсистеми дослідження поведінки споживача: у вигляді налагоджуваного медіатора (НМ₀). До складу цього пакету входять також: компонент СОМД, ІБМДМ, СВМД і ССМД; IServer₁, що надає свій інтерфейс клієнтові – додатку синхронізації даних модуля верхнього рівня. Весь цей пакет виступає у ролі моделі каталогу проектних патернів по відношенню до модуля наступного рівня.

2) Модуль ринкового рівня, або сервісний центр по регіонах (захід, схід, південь, північ, зарубіжжя), об'єднує у собі модуль дослідження ринку і товару, модуль управління рекламною діяльністю. До складу пакету цього рівня входить такий додатковий компонент як МОЗФМ, який забезпечує створення групи користувачів даного модуля, а його сервіс синхронізації маркетингових даних за допомогою IServer також виконує роль інтерфейсу клієнта, але вже по відношенню до модуля наступного, маркетингового рівня.

3) Модуль маркетингового рівня складається з таких підсистем: ціноутворення, розробка стратегії і планування діяльності підприємства, контроль маркетинговою діяльністю системи; він є пакетом компонентів, які забезпечують функціональність WEB-ПК (див. рис. 1).

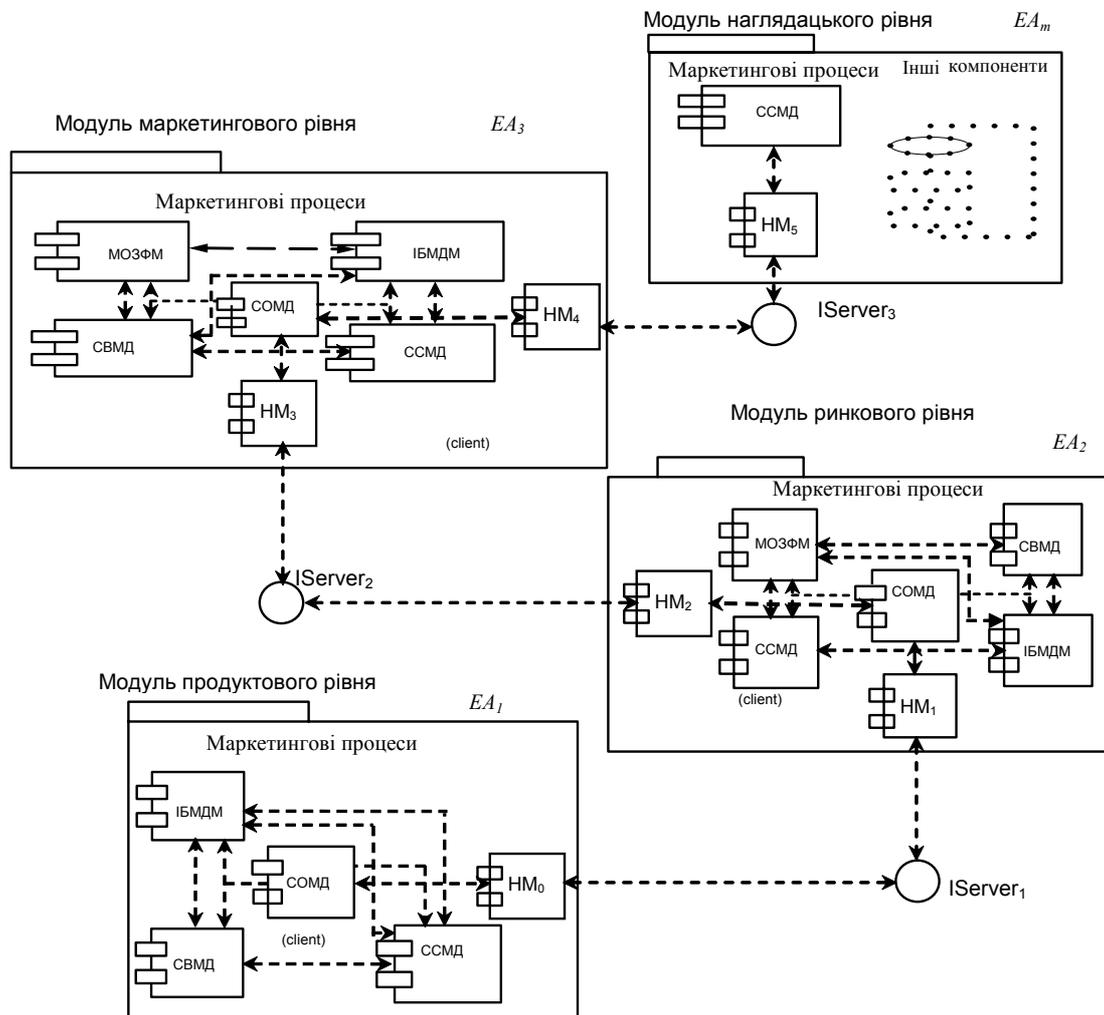


Рис. 1. Компонентна діаграма проектних шаблонів розподіленої МІС

Отже, у результаті узагальнення проектного досвіду і застосування евристичних методів ідентифікації та відбору проектних шаблонів (логічні макро-оператори IDENTIFY і SELECT, які були детально розглянуті у [6]) сформовано модель комплексу проектних патернів (ПрП) у вигляді кортежу:

$$\text{ПрП} = \langle \text{НМ}_1, \text{СОМД}, \text{ССМД}, \text{СВМД}, \text{ІБМДМ}, \text{МОЗФМ}, \text{IServer} \rangle, \quad (1)$$

де до вже введених до розгляду проектних шаблонів 1) - 5) доданий ще один: компонент налагоджуваний медіатор (НМ), що є типовим проектним рішенням для відокремлених модулів МІС [4].

Сформований таким чином каталог патернів дозволяє у компактній символічній формі, використовуючи рекурсивні вирази, показати, що запропонована модель еталонної архітектури дійсно є відкритою і масштабованою, оскільки:

1) для модуля базового рівня модель комплексу проектних патернів еталонної архітектури EA_1 може бути визначена як (див. рис. 1):

$$EA_1 = \langle (\text{НМ}_0 \cap (\text{СОМД} \cup \text{ССМД} \cup \text{СВМД} \cup \text{ІБМДМ})) \cup \text{IServer}_1 \rangle; \quad (2)$$

2) для моделі комплексу проектних патернів еталонної архітектури модуль ринкового рівня – об'єкта EA_2 можна записати у вигляді кортежу:

$$EA_2 = \langle (NM_1 \cap (CCMD \cup CBMD \cup IBMDV \cup MOZFM)_2 \cap NM_2) \cup IServer_2 \rangle ; \quad (3)$$

3) аналогічно, для моделі комплексу проектних патернів еталонної архітектури EA_3 модуля маркетингового рівня формальний опис має вигляд кортежу:

$$EA_3 = \langle (NM_2 \cap (CCMD \cup CBMD \cup IBMDV \cup MOZFM)_3 \cap NM_3) \cup IServer_3 \rangle ; \quad (4)$$

4) у загальному випадку, для модуля на рівні моделі комплексу проектних патернів еталонної ієрархії m буде справедливе визначення його еталонної архітектури EA_m у вигляді:

$$EA_m = \langle (NM_{m-1} \cap (CCMD \cup \dots)_m \cap NM_m) \cup IServer_m \rangle . \quad (5)$$

Таким чином, система моделей (2 - 5) описує процес проектування з різноманітних рівнів подання розподіленої МІС і тим самим підтверджує концептуальні положення про можливість розробки формалізованих процедур синтезу і подальшого застосування ЕА і ПР для об'єктів маркетингу різного типу, які потрібно автоматизувати на підприємстві [6].

Каталог типових архітектурних рішень, які запропоновані вище (1 - 5), повністю забезпечує реалізацію функціональності розподіленої маркетингової інформаційної системи, що розробляється, для промислових підприємств і може служити одним з варіантів ЕА для Web-базованої багаторівневої розподіленої МІС.

Для реалізації цих проектних рішень, у ході узагальнення результатів виконання ряду конкретних проектів з розробки і експлуатації маркетингових інформаційних систем різного рівня для ТОВ «АВМ Ампер», із застосуванням загальної концептуальної процедури розробки колекції патернів, розробленої у [4], були отримані шаблони для створення повторно-використовуваних програмних компонентних рішень. Як приклад використання запропонованої моделі каталогу проектних компонентів наведемо опис одного з найбільш важливих отриманих при цьому елементів, тобто проектного шаблону для сервісу обміну маркетинговими даними.

СОМД - це програмний сервіс, який повинен підтримувати у МІС коректний обмін маркетинговими даними з архітектурними модулями системи на кожному топологічному рівні (наприклад: модуль дослідження поведінки споживачів (продуктовий рівень), модуль дослідження ринку і товарів (ринковий рівень), модуль ціноутворення або управління рекламною діяльністю (маркетинговий рівень)), котрі входять до функціональної структури відповідної МІС. Крім того, СОМД надає свої інтерфейси клієнтським додаткам, які потім виконують такі процедури як візуалізація цих даних, побудова графіків різних трендів і сегментів ринку, позиціонування продукції та споживачів, відповідно до загальної функціональності МІС.

Для реалізації СОМД пропонуємо створити компонентну структуру, котра повинна забезпечити одночасну взаємодію декількох клієнтських додатків з декількома (і, як правило, різнотипними) зовнішніми медіаторами [2, 4]. Для цього у структурі СОМД доцільно виділити два компоненти (рис. 2): XProtocol і TechXObject.

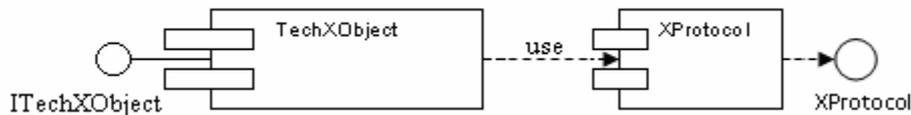


Рис. 2. Компонентна діаграма сервісу обміну маркетинговими даними

XProtocol – це компонент СОМД, який реалізує отримання маркетингових даних від медіаторів, посилання запитів і синхронізацію отриманої маркетингової інформації, щоб забезпечити цілісність даних;

TechXObject – компонент СОМД, який забезпечує надання маркетингових даних іншим додаткам-клієнтам (далі – просто клієнтам) за допомогою реалізації інтерфейсу ITechXObject.

Отже, TechXObject є менеджером даних у складі СОМД і він відповідальний за виконання таких задач:

- отримання клієнтського запиту;
- пошук запитаних маркетингових даних у своєму буфері пам'яті;
- синхронізація потоків компоненту XProtocol, що записують маркетингову інформацію у буфер, з потоками, які прочитують інформацію з буфера.

Порядок дій обох компонентів СОМД представлений діаграмою послідовності (рис. 3).

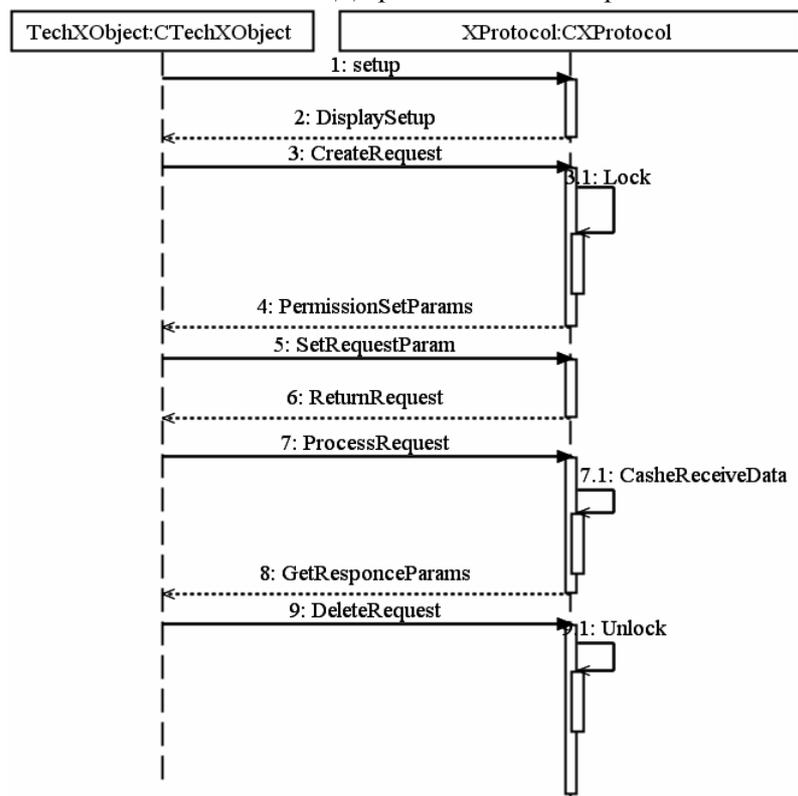


Рис. 3. Діаграма послідовності роботи компонентів сервісу обміну маркетинговими даними

При цьому виникає проблема синхронізації доступу різних клієнтів на читання і потік оновлення, на запис нових маркетингових даних, тобто поки дані, що пишуться, та потік даних, що читається, не повинен мати доступ до тих даних, які записуються. Якщо виникає така ситуація, то вона веде до взаємоблокування ділянки пам'яті загального доступу і, як наслідок, до критичної помилки у роботі всього додатку у цілому.

Найбільш очевидний вихід з цієї ситуації – використання нульової змінної, яка своїм становищем показуватиме, зайнятий об'єкт маркетингу чи ні. Це буде працездатним рішенням, але процес оновлення інформації постійно намагатиметься провести запис нових маркетингових даних, тим самим витрачаючи багато процесорного часу. Тому для рішення проблеми синхронізації слід скористатися відповідними можливостями MS Win32 API, який підтримує декілька так названих об'єктів синхронізації [9]:

м'ютекс – це об'єкт, який працює як шлюз, через який може пройти тільки один потік в один і той же час;

семафор – це об'єкт, котрий працює як шлюз, через який може проходити декілька потоків в один і той же час;

подія – це об'єкт, який генерує сигнал для усіх потоків, що слухають його;

критична секція – об'єкт, який працює як м'ютекс, але тільки у межах одного процесу.

Кожен об'єкт синхронізації координує різні типи взаємодії потоків, але всі вони працюють поодиноці за тим же принципом. Коли деякий потік повинен виконати певні скоординовані дії, то він чекає відповіді від одного з подібних об'єктів і виконує їх тільки після того, як отримає відповідь. При очікуванні потоком даних відповіді система не обробляє його, таким чином розвантажуючи процесор, а коли приходить відповідь, то потік продовжує обробку отриманих раніше маркетингових даних. Якщо декілька потоків повинні одночасно звертатися до об'єкта загального доступу, то найдоцільніше використовувати м'ютекс для синхронізації їх

дій. Так, перш ніж почати запис, потік оновлення заздалегідь запрошує відповідний м'ютекс і якщо м'ютекс вільний, то проводить запис нових маркетингових даних, інакше потік блокується і припиняється його виконання.

Результат застосування патерну.

Запропоноване рішення для СОМД забезпечує такі його характеристики:

1) конфігурованість – можливість його налаштування на будь-який заданий протокол обміну маркетинговими даними з відповідним модулем МІС;

2) продуктивність – тобто достатню швидкість роботи СОМД з прийому і обробки маркетингових даних одночасно від декількох модулів МІС.

Таким чином, чим швидше відбуватиметься обробка запитів, тим швидше відбуватиметься оновлення маркетингових даних, які потім використовуються у клієнтських додатках. Ця обставина істотна у роботі будь-якої МІС, оскільки чим вища актуальність даних, тим більш адекватні рішення з управління об'єктом маркетингу можуть бути прийняті маркетологом.

3. Висновки

Проведений аналіз отриманих архітектурних і програмних рішень показує їх складність і тим самим визначає необхідність розробки моделей та інструментальних засобів для проектування складних розподілених МІС.

Науковою новизною є удосконалена модель компонентної діаграми проектних шаблонів розподіленої МІС, подальший розвиток математичної моделі комплексу проектних патернів системи, яка на відміну від інших дозволяє моделювати архітектуру МІС залежно від автоматизації об'єктів маркетингу на підприємстві. Крім того, удосконалена процедура отримання проектного шаблону сервісу обміну маркетинговими даними, який дозволяє коректно здійснювати обмін даними з архітектурними модулями системи на кожному топологічному рівні МІС, що дозволяє швидше оброблювати запити та оновлення маркетингових даних, оскільки чим вища актуальність даних, що відображаються у системі, тим більш адекватні рішення з управління маркетинговими процесами приймаються менеджером з маркетингу.

Практичним значенням є проектні рішення: інтегрована база маркетингових даних модуля; сервіс візуалізації маркетингових даних; сервіс синхронізації маркетингових даних; менеджер облікових записів фахівців-маркетологів, що надає можливість створювати типові комплекси проектних патернів еталонних архітектур маркетингових інформаційних систем у відповідності з рішенням маркетингових задач на підприємствах різних галузей. Це скорочує час проектування МІС, зменшує матеріальні і фінансові витрати.

Список літератури: 1. *Ткачук М.В.* Про один підхід до побудови архітектури розподіленої обробки даних в інтранет-системах управління технологічними процесами / М.В. Ткачук, Д.В. Кукленко // Системні дослідження та інформаційні технології. 2003. № 4. С. 35-43. 2. *Костенко О.П.* Моделирование функциональной структуры информационно-аналитической маркетинговой системы // Нові технології. Науковий вісник ІЕНТ. 2004. № 3 (6). С. 127-130. 3. *Левикін В.М.* Розробка просторово-траєкторного підходу для проектування маркетингових інформаційних систем / Левикін В.М., Костенко О.П. // Сьома дистанційна науково-практична конференція за міжнародною участю «системи підтримки прийняття рішень. Теорія і практика. СППР-2011». Червень, 2011 р., Україна, Київ. С 50 – 53. 4. *Левикін В.М.* Розробка концепції інтегрованого модельно-технологічного інструментарію для проектування МІС / В.М. Левикін, О.П. Костенко, В.В. Хміль-Чуприна // Нові технології. Науковий вісник КУЕІТУ. 2011. № 1(31). С. 58-62. 5. *Буч Г.* UML – руководство пользователя. / Г. Буч, Д. Рамбо, А. Джекобсон. // М.: ДМК Пресс, 2001. 432 с. 6. *Левикін В.М.* Розробка математичної моделі еталонної системної архітектури маркетингових ІС / В.М. Левикін, О.П. Костенко, В.В. Хміль-Чуприна // Перспективи інформаційних технологій. Херсон -2011. № 2 (008). С. 41-47. 7. *Архитектуры, модели и технологии программного обеспечения информационно-управляющих систем: Монография / Н.В. Ткачук, В.А. Шеховцов, Д.В. Кукленко, В.Е. Сокол / Под ред. М.Д. Годлевского. Харьков: НТУ «ХПИ», 2005. 546 с.*

Надійшла до редколегії 20.05.2011

Левикін Віктор Макарович, д-р техн. наук, професор, завідувач кафедри ІУС ХНУРЕ. Наукові інтереси: розробка розподілених інформаційних систем. Адреса: Україна, 61166, Харків, пр. Леніна, 14, тел. 70-21-451.

Костенко Олександр Петрович, канд. техн. наук, доцент кафедри інформаційно-управляючих систем, Кременчуцький національний університет ім. М. Остроградського. Наукові інтереси: розробка маркетингових інформаційних систем. Адреса: Україна, 39600, Кременчук, вул. Першотравнева, 20, тел. 0674176150. e-mail: kap_ua@rambler.ru.

Зінченко Євгеній Григорович, старший викладач кафедри інформатики Кременчуцького університету економіки, інформаційних технологій і управління. Наукові інтереси: розробка складних інформаційних систем. Адреса: Україна, 39600, Кременчук, вул. Пролетарська 24/37.

УДК 681.3

Ю.В. УЛЬЯНОВСКАЯ

ИССЛЕДОВАНИЕ СВОЙСТВ ОТНОШЕНИЙ НЕЧЕТКОЙ БЛИЗОСТИ ОБЪЕКТОВ В ИНФОРМАЦИОННЫХ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ

Рассматриваются нечеткие отношения близости объектов на основе нечеткой близости, нечеткого включения и нечеткой общности ситуаций для проведения идентификации в интеллектуальных системах с учетом возможности изменения информации об объектах во времени. Для каждого отношения исследуются его свойства. На основе проведенных исследований делается вывод о типе отношения.

1. Актуальность задачи

В основе работ большинства систем искусственного интеллекта заложены знания экспертов, которые формализуются с помощью методов инженерии знаний. При первоначальном формировании базы знаний очень важным является вопрос о методах получения и обработки экспертной информации, характер которой влияет на выбор модели представления знаний, метода обработки и формализации знаний и на структуру базы знаний в целом. Процесс получения знаний от экспертов включает несколько этапов: подбор экспертов, их опрос, обработка экспертных оценок. Экспертную оценку объектов можно рассматривать как процесс получения экспертной информации с помощью измерений. Сами объекты при этом могут быть описаны с помощью как качественных, так и количественных данных. Для принятия решений в системах искусственного интеллекта при идентификации объектов важно не только сравнивать их между собой, но и знать, во сколько раз и на сколько условных единиц один объект предпочтительнее другого [1]. Решение этой задачи выполняется путем применения того или иного метода получения качественных или количественных оценок. При этом необходимо не только получить экспертную оценку, но и провести ранжирование альтернатив, разбиение их на классы эквивалентности, определить сравнительную предпочтительность альтернатив путем построения отношений на их множестве. Это порождает задачу проверки таких свойств как транзитивность, симметричность и других свойств отношений. В зависимости от этого на множестве альтернатив могут быть заданы отношения различных типов [2].

2. Основные направления исследований

Решение сформулированной выше задачи происходит в нескольких направлениях. Экспертная информация о большинстве предметных областей является совокупностью качественных и количественных оценок и характеризуется неполнотой и нечеткостью. Если оценки эксперта носят качественный характер, то для описания связей применяются отношения линейного или частичного порядка, эквивалентности, толерантности, а иногда и произвольные отношения, не обладающие такими свойствами, как связность, транзитивность и т.д. Если экспертная информация содержит количественные оценки, используются метризованные отношения соответствующего типа [2]. В работе [3] рассматриваются

методы обработки экспертной информации для определения степени близости объектов в зависимости от ее типа. Для количественной экспертной информации проведен анализ мер близости в зависимости от типа шкалы, в которой проводились измерения. Для качественной информации проведен анализ степеней нечеткой близости ситуаций и построены отношения нечеткого включения, нечеткого равенства и нечеткой общности с учетом изменения информации.

Целью данной работы является исследование свойств отношений нечеткой близости объектов, а именно нечеткого включения, нечеткого равенства и нечеткой общности с учетом изменения информации об объектах во времени и определение типов указанных отношений.

3. Исследование свойств отношений нечеткой близости объектов

Предметные области, с которыми работают экспертные системы, весьма разнообразны. Полностью описать факты, правила и взаимосвязи в предметной области не всегда возможно. Это порождает проблему неполноты и неточности данных. При использовании в описании качественных характеристик возникает необходимость в обработке нечетких данных с применением лингвистических переменных. Решение сформулированных задач возможно с помощью аппарата нечетких множеств. Пусть x – объект, оцениваемый экспертами, описывается совокупностью признаков $\Xi = \{\xi_1, \xi_2, \dots, \xi_n\}$. Пусть признаки $\xi_i \in \Xi$ относятся к одному из трех типов: числовому, булевому или лингвистическому. Каждый лингвистический атрибут $\xi_i, (i \in I = \{1, \dots, N\})$ описывается соответствующей лингвистической переменной ξ_i, Ξ_i, D_i , где $\Xi_i = \{\xi_1^i, \xi_2^i, \dots, \xi_n^i\}$ – терм-множество лингвистической переменной ξ_i (набор лингвистических значений признака), m_i – число значений признака, D_i – базовое множество признака ξ_i . Для описания термов $\xi_i^j (j \in L = \{1, 2, \dots, m_i\})$, соответствующих значениям признака ξ_i , используются нечеткие переменные $\langle \xi_i^j, D_i, \tilde{C}_i^j \rangle$, т.е. значение ξ_i^j описывается нечетким множеством \tilde{C}_i^j в базовом множестве D_i : $\tilde{C}_i^j = \{\langle \mu_{\tilde{C}_i^j}(d)/d \rangle\}, d \in D_i$. Таким образом, в терминах теории нечетких множеств каждый объект $x \in X$ можно представить в виде:

$$x = \{\langle \mu_x(\xi_i)/\xi_i \rangle\}, \xi_i \in \Xi, \quad (1)$$

где $\mu_x(\xi_i) = \{\langle \mu_{\mu_x(\xi_i)}(\xi_i^j)/\xi_i^j \rangle\}, j \in L, i \in I$. Для признаков ξ_i булевого типа $\mu_x(\xi_i) \in \{0;1\}$.

Признаки ξ_i могут изменяться во времени постепенно либо скачкообразно [3]. Учитывая это, введем коэффициент t_i , характеризующий изменения во времени значения ξ_i -го признака. В связи с этим каждый признак охарактеризуем следующими параметрами: $\mu_x(\xi_i^j)$ – функция принадлежности, которая показывает, в какой мере значение ξ_i^j признака ξ_i характеризует объект x ; t_i – коэффициент изменения во времени значения ξ_i -го признака.

Как отмечалось выше, одной из задач, которые решают экспертные системы, является задача идентификации, при решении которой необходимо определить степень близости объектов. Пусть системе необходимо идентифицировать объект y через определение степени его близости к объекту x , хранящемуся в базе знаний экспертной системы. Предположим, что все признаки, которые характеризуют идентифицируемые объекты x и y , являются хорошо определенными. Признак ξ_i считается плохо определенным для x , если $\exists \xi_i^j \in \Xi_i : \mu_{\mu_x(\xi_i)}(\xi_i^j) \in (1 - T, T)$, где T – некоторый порог значимости [2]. Данное утверждение имеет смысл при $T \geq 0,5$.

Результатом сравнения объектов между собой может быть их полное или частичное совпадение. В зависимости от количества совпадающих признаков между объектами выполняется отношение нечеткого равенства, нечеткого включения и нечеткой общности. Учитывая, что признаки, на основании которых проводятся сравнения объектов, могут

изменяться во времени, достоверность идентификации снижается. Может возникнуть две ситуации. В первом случае значение всех признаков определены, при этом они могут совпадать либо полностью, либо частично. Во втором случае часть признаков может быть не определена. Рассмотрим первый случай. Степень нечеткой близости объектов определим через степень нечеткого равенства $\mu(x, y)$, при этом равенство ξ_i признака определим через степень равенства его значений:

$$\mu(x, y) = \mu(\mu_{\bar{x}}(\xi_i), \mu_{\bar{y}}(\xi_i)) = \&_{\xi_j^i \in \xi_i} (\mu_{\mu_x(\xi_i)}(\xi_j^i) \leftrightarrow \mu_{\mu_y(\xi_i)}(\xi_j^i)). \quad (2)$$

Учитывая коэффициент изменения во времени t_i , степень нечеткого равенства объектов вычислим следующим образом [3]:

$$\mu(x, y) = \&_{\xi_i \in \Xi} \mu(\mu_{\bar{x}}(\xi_i), \mu_{\bar{y}}(\xi_i)) = \&_{\xi_i \in \Xi} \mu(\min\{1 - t_i, \mu_{\bar{x}}(\xi_i)\}, \mu_{\bar{y}}(\xi_i)). \quad (3)$$

Если $\exists \xi_i : 0 < t_i$, то степень достоверности определения меры сходства принимается равной минимальной степени достоверности равенства изменяющихся признаков, которая в свою очередь определяется при помощи коэффициента t_i изменения признака ξ_i во времени.

Определяя таким образом степень равенства ситуаций x и y , ненулевое значение функции принадлежности $\mu(x, y)$ получим в том случае, когда будут совпадать все значения атрибутов булевого типа.

Пусть при сравнении объектов x и y часть признаков не равны между собой. Через q обозначим количество нечетко неравных признаков. В этом случае степень нечеткой близости определим через нечеткую $(n-q)$ общность $k_{n-q}(x, y)$ следующим образом:

$$k_{n-q}(x, y) = \&_{\xi_i \in \Xi \setminus \Xi^Q} \mu(\min\{1 - t_i, \mu_{\bar{x}}(\xi_i)\}, \mu_{\bar{y}}(\xi_i)), \quad (4)$$

$|\Xi^Q| \leq q$, $\xi_k \in \Xi^Q \Leftrightarrow \mu(\mu_x(\xi_k), \mu_y(\xi_k)) \leq R$, где R – некоторое пороговое, наперед заданное значение: $R \in [0; 1]$.

При определении $(n-q)$ общности x и y не учитываем признаки, значения которых нечетко не равны для x и y . При $\Xi^Q = \emptyset$ объекты нечетко равны.

Рассмотрим второй случай. Пусть значения части признаков объекта y не определены. Тогда число признаков, описывающих x , больше, чем число признаков, характеризующих y . Воспользуемся формулой нечеткого включения:

$$v(y, x) = \&_{\xi_i \in \Xi} v(\mu_{\bar{y}}(\xi_i), \mu_{\bar{x}}(\xi_i)), \quad (5)$$

где $v(\mu_{\bar{y}}(\xi_i), \mu_{\bar{x}}(\xi_i)) = \&_{\xi_j^i \in \xi_i} (\mu_{\mu_y(\xi_i)}(\xi_j^i) \rightarrow \mu_{\mu_x(\xi_i)}(\xi_j^i))$ и является степенью включения нечетного множества $\mu_y(\xi_i)$ в нечетное множество $\mu_x(\xi_i)$.

Учитывая изменение признака во времени, формула (5) принимает вид

$$v(y, x) = \&_{\xi_i \in \Xi} v(\mu_{\bar{y}}(\xi_i), \min\{1 - t_i, \mu_{\bar{x}}(\xi_i)\}). \quad (6)$$

Будем предполагать, что объекты нечетко близки, если соответствующая мера близости больше некоторого порогового значения T . Введенные выше меры близости для объектов с изменяющимися признаками могут быть рассмотрены как нечеткие отношения. Как и для четких отношений, тип нечеткого отношения определяется совокупностью его свойств. К основным свойствам отношений относятся свойства рефлексивности, антирефлексивности, симметричности, несимметричности, антисимметричности, транзитивности, связности. Основными типами отношений являются отношения порядка, толерантности, эквивалентности, доминирования. Для нечетких отношений вводится понятие степени, отражающее их нечеткость.

Пусть дано произвольное нечеткое отношение $\tilde{\varphi} = (X, \tilde{F})$. В работе [2] вводятся следующие определения степени отношений.

Степенью рефлексивности $\alpha(\tilde{\varphi})_{\text{ref}}$ называется величина, определяемая выражением

$$\alpha(\tilde{\varphi})_{\text{ref}} = \&_{x \in X} \mu_F \langle x, x \rangle. \quad (7)$$

Отношение $\tilde{\varphi} = (X, \tilde{F})$ называется нечетко рефлексивным, если $\alpha(\tilde{\varphi})_{\text{ref}} \geq 0,5$, нечетко нереплексивным, если $\alpha(\tilde{\varphi})_{\text{ref}} \leq 0,5$. Если $\alpha(\tilde{\varphi})_{\text{ref}} = 0,5$, то отношение $\tilde{\varphi}$ называется рефлексивно индифферентным.

Степенью антирефлексивности называется величина

$$\beta(\tilde{\varphi})_{\text{ref}} = \&_{x \in X} (\neg \mu_F \langle x, x \rangle) = \neg \left(\vee_{x \in X} \mu_F \langle x, x \rangle \right). \quad (8)$$

Отношение $\tilde{\varphi}$ называется нечетко антирефлексивным, если $\beta(\tilde{\varphi})_{\text{ref}} \geq 0,5$, и нечетко неантирефлексивным, если $\beta(\tilde{\varphi})_{\text{ref}} \leq 0,5$. При $\beta(\tilde{\varphi})_{\text{ref}} = 0,5$ оно антирефлексивно индифферентно.

Степенью симметричности $\alpha(\tilde{\varphi})_{\text{sym}}$ называется величина

$$\alpha(\tilde{\varphi})_{\text{sym}} = \&_{\substack{x \in X \\ x \neq y}} (\mu_F \langle x, y \rangle \leftrightarrow \mu_F \langle y, x \rangle).$$

Отношение $\tilde{\varphi}$ называется нечетко симметричным, если $\alpha(\tilde{\varphi})_{\text{sym}} \geq 0,5$, и нечетко несимметричным, если $\alpha(\tilde{\varphi})_{\text{sym}} \leq 0,5$. При $\alpha(\tilde{\varphi})_{\text{sym}} = 0,5$ – симметрично индифферентным.

Степенью антисимметричности $\beta(\tilde{\varphi})_{\text{sym}}$ называется величина

$$\beta(\tilde{\varphi})_{\text{sym}} = \&_{\substack{x, y \in X \\ x \neq y}} \neg (\mu_F \langle x, y \rangle \& \mu_F \langle y, x \rangle). \quad (9)$$

Отношение $\tilde{\varphi}$ называется нечетко антисимметричным, если $\beta(\tilde{\varphi})_{\text{sym}} \geq 0,5$, и нечетко неасимметричным, если $\beta(\tilde{\varphi})_{\text{sym}} \leq 0,5$. В случае, если $\beta(\tilde{\varphi})_{\text{sym}} = 0,5$, отношение $\tilde{\varphi}$ называется антисимметрично индифферентным.

Степенью транзитивности $\alpha(\tilde{\varphi})_{\text{tr}}$ отношения $\tilde{\varphi}$ называется величина

$$\alpha(\tilde{\varphi})_{\text{tr}} = \&_{\substack{x, y, z \in X \\ x \neq y \neq z}} \left(\left(\vee_y (\mu_F \langle x, y \rangle \& \mu_F \langle y, z \rangle) \rightarrow \mu_F \langle x, z \rangle \right) \right) \quad (10)$$

Отношение $\tilde{\varphi}$ называется нечетко транзитивным, если $\alpha(\tilde{\varphi})_{\text{tr}} \geq 0,5$, нечетко нетранзитивным, если $\alpha(\tilde{\varphi})_{\text{tr}} \leq 0,5$, транзитивно индифферентным, если $\alpha(\tilde{\varphi})_{\text{tr}} = 0,5$.

Пусть $X = \{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_N\}$ есть некоторое множество типовых ситуаций. Покажем, что отношение нечеткого включения $\tilde{\delta} = (X, \tilde{F})$ является отношением нечеткого нестрогого порядка, т.е. является рефлексивным, антисимметричным и транзитивным. При этом $\tilde{F} = \left\langle \mu_F \langle \tilde{x}_i, \tilde{x}_j \rangle / \langle \tilde{x}_i, \tilde{x}_j \rangle \right\rangle$ является отношением нечеткого включения, $\mu_F \langle \tilde{x}_i, \tilde{x}_j \rangle = v \langle \tilde{x}_i, \tilde{x}_j \rangle$, $v \langle \tilde{x}_i, \tilde{x}_j \rangle$ определяется выражением (6).

Для доказательства достаточно показать, что $\alpha(\tilde{\delta})_{\text{ref}} \& \beta(\tilde{\delta})_{\text{sym}} \& \alpha(\tilde{\delta})_{\text{tr}} \geq 0,5$ или, с учетом вида отношения $\tilde{\delta}$

$$\alpha(\tilde{\delta})_{\text{ref}} \& \beta(\tilde{\delta})_{\text{sym}} \& \alpha(\tilde{\delta})_{\text{tr}} \geq \Gamma. \quad (11)$$

Покажем, что $\alpha(\tilde{\delta})_{\text{ref}} \geq \Gamma$. С учетом выражений (5) и (7) необходимо показать, что $\forall \tilde{x} \in X$ справедливо $v(\tilde{x}, \tilde{x}) \geq \Gamma$. С учетом выражения способа определения нечеткого включения можем записать: $v(\tilde{x}, \tilde{x}) = \&_{\xi_i \in \Xi} v(\mu_x(\xi_i), \min\{1 - t_i, \mu_x(\xi_i)\})$.

Необходимо показать, что $\bigwedge_{\xi_i \in \Xi} v(\mu_x(\xi_i), \mu_x(\xi_i)) \geq T$. С учетом (1) и (2) требуется показать, что $\forall \xi_j^i \in \xi_i$:

$$\neg \mu_{\mu_x(\xi_i)}(\xi_j^i) \vee \min\{1 - t_i, \mu_{\mu_x(\xi_i)}(\xi_j^i)\} \geq T. \quad (12)$$

Пусть $t_i \geq T$, тогда $1 - t_i \leq T$. По условию хорошо определенной ситуации $\mu_{\mu_x(\xi_i)}(\xi_j^i) < 1 - T$, или $\mu_{\mu_x(\xi_i)}(\xi_j^i) > T$. Если $\mu_{\mu_x(\xi_i)}(\xi_j^i) < 1 - T$, то $\neg \mu_{\mu_x(\xi_i)}(\xi_j^i) > T$ и неравенство (12) выполняется. Если $\mu_{\mu_x(\xi_i)}(\xi_j^i) > T$, то $\neg \mu_{\mu_x(\xi_i)}(\xi_j^i) < T$ и неравенство не выполняется.

Пусть $t_i \leq T$, тогда $1 - t_i \geq T$. Если $\mu_{\mu_x(\xi_i)}(\xi_j^i) < 1 - T$, то $\neg \mu_{\mu_x(\xi_i)}(\xi_j^i) > T$ и неравенство (12) выполняется благодаря левой части дизъюнкции. Если $\mu_{\mu_x(\xi_i)}(\xi_j^i) > T$, то $\neg \mu_{\mu_x(\xi_i)}(\xi_j^i) < 1 - T$ а $\min\{1 - t_i, \mu_{\mu_x(\xi_i)}(\xi_j^i)\} \geq T$ и неравенство выполняется благодаря правой части. Таким образом, неравенство (10) выполняется, когда $t_i \leq T$ при любом значении $\mu_x(\xi_i)$, либо когда $t_i \geq T$ и $\mu_x(\xi_i) \leq 1 - T$. Неравенство $\alpha(\tilde{\delta})_{\text{ref}} \geq T$ доказано. Тем самым доказано, что отношение нечеткого включения является рефлексивным при $t_i \leq T$ или при $t_i \geq T$ и $\mu_x(\xi_i) \leq 1 - T$.

Докажем, что $\beta(\tilde{\delta})_{\text{sym}} \geq T$. В соответствии с выражением (7) $\beta(\tilde{\delta})_{\text{sym}} = \bigwedge_{\substack{x_i, x_j \in X \\ x_i \neq x_j}} \neg(\mu_F(x_i, x_j) \& \mu_F(x_j, x_i))$, или $\beta(\tilde{\delta})_{\text{sym}} = \bigwedge_{\substack{x_i, x_j \in X \\ x_i \neq x_j}} \neg(v(x_i, x_j) \& v(x_j, x_i))$, т.е. не-

обходимо показать, что $\forall x_i, x_j: x_i, x_j \in X, x_i \neq x_j (\neg(v(x_i, x_j) \& v(x_j, x_i)) \geq T)$.

Неравенство $(\neg(v(x_i, x_j) \& v(x_j, x_i)) \geq T)$ эквивалентно $(v(x_i, x_j) \& v(x_j, x_i)) \leq 1 - T$. Поскольку при определении выражения (6) предполагалось, что $(\forall \tilde{x}_i, \tilde{x}_j \in X)$, имеет место утверждение $((i \neq j \& \tilde{x}_i \subseteq \tilde{x}_j) \rightarrow (\tilde{x}_i \not\subseteq \tilde{x}_j))$. Иначе \tilde{x}_i и \tilde{x}_j нужно воспринимать как одну ситуацию. Отсюда следует, что если $v(x_i, x_j) \geq T$, то $v(x_j, x_i) \leq 1 - T$, и наоборот. Следовательно, можем сделать вывод, что неравенство $(v(x_i, x_j) \& v(x_j, x_i)) \leq 1 - T$ выполняется всегда.

Докажем, что $\alpha(\tilde{\delta})_{\text{tr}} \geq T$. Аналогично проведенным выше доказательствам с учетом вида отношения $\tilde{\delta}$ необходимо показать, что $\forall x, y \in X, x \neq y, \exists z \in X: x \neq z, z \neq y$, для которого $(v(x, y) \geq T \& v(y, z) \geq T) \rightarrow v(x, z) \geq T$.

Выражение $(v(x, y) \geq T \& v(y, z) \geq T)$ эквивалентно выражению $(v(x, y) \& v(y, z)) \geq T$. С учетом (5) и (6) последнее неравенство имеет вид

$$(\mu_x(\xi_i) \rightarrow \min\{1 - t_i, \mu_y(\xi_i)\}) \& (\mu_y(\xi_i) \rightarrow \min\{1 - t_i, \mu_z(\xi_i)\}) \geq T. \quad (13)$$

Проведя рассуждения, аналогичные доказательству рефлексивности, получим, что данное неравенство, а вместе с ним и нечеткая транзитивность выполняется, если $t_i \leq T$ и $\mu_x(\xi_i) \leq 1 - T$, либо если $t_i \geq T$ и $\mu_x(\xi_i) \leq 1 - T$.

Учитывая, ограничения, при которых отношение удовлетворяет свойству рефлексивности, можем сделать вывод, что отношение нечеткого включения $\tilde{\delta} = (X, \tilde{F})$ является отношением нечеткого нестрогого порядка при $t_i \leq T$ и $\mu_x(\xi_i) \leq 1 - T$.

Нечеткое отношение $\tilde{\varphi} = (S, \tilde{F})$, где

$$\tilde{F} = \left\langle \mu_F \langle \tilde{x}_i, \tilde{x}_j \rangle / \langle \tilde{x}_i, \tilde{x}_j \rangle \right\rangle, \quad (14)$$

является отношением нечеткого равенства, если $\mu_F \langle \tilde{x}_i, \tilde{x}_j \rangle = \mu \langle \tilde{x}_i, \tilde{x}_j \rangle$.

Покажем, что отношение нечеткого равенства $\tilde{\varphi} = (X, \tilde{F})$ является отношением нечеткой эквивалентности. При этом $\tilde{F} = \left\langle \mu_F \langle \tilde{x}_i, \tilde{x}_j \rangle / \langle \tilde{x}_i, \tilde{x}_j \rangle \right\rangle$ является отношением нечеткого равенства, если $\mu_F \langle \tilde{x}_i, \tilde{x}_j \rangle = \mu \langle \tilde{x}_i, \tilde{x}_j \rangle$, $\mu \langle \tilde{x}_i, \tilde{x}_j \rangle$ определяется выражениями (2),(3).

Отношение называется отношением эквивалентности, если оно рефлексивно, симметрично и транзитивно. Для доказательства нечеткой эквивалентности необходимо доказать, что $\alpha(\tilde{\varphi})_{ref} \& \alpha(\tilde{\varphi})_{sym} \& \alpha(\tilde{\varphi})_{tr} \geq 0,5$ или, с учетом вида отношения $\tilde{\varphi}$

$$\alpha(\tilde{\varphi})_{ref} \& \alpha(\tilde{\varphi})_{sym} \& \alpha(\tilde{\varphi})_{tr} \geq T. \quad (15)$$

Покажем, что $\alpha(\tilde{\varphi})_{ref} \geq T$. С учетом выражений (7) и вида отношения $\tilde{\varphi}$ можем записать, что $\alpha(\tilde{\varphi})_{ref} = \&_{\tilde{x} \in X} \mu(x, x)$. Необходимо показать, что $\forall \tilde{x} \in X$ справедливо $\mu(\tilde{x}, \tilde{x}) \geq T$. С учетом способа определения нечеткого равенства $\mu(\tilde{x}, \tilde{x}) = \& \mu(\min\{(1-t_i), \mu_x(\xi_i)\}, \mu_x(\xi_i))$ необходимо показать $\&_{\xi_i \in \Xi} \mu(\min\{(1-t_i), \mu_x(\xi_i)\}, \mu_x(\xi_i)) \geq T$. Это эквивалентно утверждению, что $(\forall \xi_j^i \in \xi_i)$:

$$(\neg \min((1-t_i), \mu_{\mu_x(\xi_i)}(\xi_j^i)) \vee \mu_{\mu_x(\xi_i)}(\xi_j^i)) \wedge (\min((1-t_i), \mu_{\mu_x(\xi_i)}(\xi_j^i)) \vee \neg \mu_{\mu_x(\xi_i)}(\xi_j^i)) \geq T. \quad (16)$$

Неравенство (16) выполняется при $\mu_{\mu_x(\xi_i)}(\xi_j^i) \leq 1-T$. Таким образом, $\alpha(\tilde{\delta})_{ref} \geq T$ при любом t_i и $\mu_{\mu_x(\xi_i)}(\xi_j^i) \leq 1-T$.

Проверим выполнение неравенства $\alpha(\tilde{\varphi})_{sym} \geq T$. Учитывая (10), имеем

$$\alpha(\tilde{\varphi})_{sym} = \&_{\substack{x, y \in X \\ x \neq y}} (\neg \mu_F \langle x, y \rangle \vee \mu_F \langle y, x \rangle) = \&_{\substack{x, y \in X \\ x \neq y}} (\neg \mu(x, y) \vee \mu(y, x))$$

Необходимо, чтобы $\neg \mu(x, y) \vee \mu(y, x) \geq T$. Последнее неравенство выполняется для любых t_i .

Аналогично доказывается, что $\alpha(\tilde{\delta})_{tr} \geq T$. Это доказывает, что отношение (2), где μ определяется выражением (3), является отношением эквивалентности при ограничениях на t_i . Необходимо отметить, что если $T=0,5$, то рассмотренные выше ограничения на t_i снимаются.

Рассмотрим отношение нечеткой общности ситуаций. Нечеткое отношение называется отношением $(n-q)$ - общности, если $\mu_{\tilde{F}} \langle \tilde{x}_i, \tilde{x}_j \rangle = k_{n-q}(\tilde{x}_i, \tilde{x}_j)$, где $\tilde{x}_i, \tilde{x}_j \in X$. Поскольку $\forall \tilde{x}_i, \tilde{x}_j \in X : \mu \langle \tilde{x}_i, \tilde{x}_j \rangle \leq k_{n-q}(\tilde{x}_i, \tilde{x}_j)$, то $\alpha(\tilde{\tau})_{ref} \geq \alpha(\tilde{\varphi})_{ref}$ и $\alpha(\tilde{\tau})_{sym} \geq \alpha(\tilde{\varphi})_{sym}$, где $\tilde{\varphi}$ -отношение нечеткого равенства, можем сделать вывод, что $\alpha(\tilde{\tau})_{ref} \geq T$, $\alpha(\tilde{\tau})_{sym} \geq T$. Это в свою очередь обозначает, что отношение $\tau = (S, \tilde{F})$ является нечетко рефлексивным и нечетко симметричным, степень толерантности $\alpha(\tilde{\tau})_{ref} \& \alpha(\tilde{\tau})_{sym} \geq T$ и $\tau = (S, \tilde{F})$ является отношением нечеткой толерантности.

Выводы

Проанализированы свойства отношений нечеткой близости объектов. Показано, что при определенных ограничениях на параметр t_i отношение нечеткого включения является отношением нечеткого нестрогого порядка. Это дает возможность решить поставленную задачу идентификации объектов путем организации на множестве X иерархии типовых

объектов. Для разбиения множества X на классы нечеткой эквивалентности необходимо применить к объектам множества X отношение нечеткого равенства $\tilde{\varphi} = (X, \tilde{F})$, которое, как было показано в работе, является отношением нечеткой эквивалентности при ограничениях на t_i . При этом в один класс будут входить нечетко равные между собой объекты, которые, с учетом порога T , можно считать одним объектом. Для полного описания объектов необходимо не только знать их признаки, но и выбрать соответствующую шкалу для их измерения. Поскольку между основными типами шкал и отношениями существует непосредственная связь, перспективным является построение шкал с учетом типов отношений в рассматриваемой предметной области.

Список литературы: 1. Ульяновська Ю. Аналіз основних аспектів побудови інтелектуальної автоматизованої системи ідентифікації творів мистецтва // Вісник Академії митної служби України. 2002. №1. С. 70-74. 2. Литвак Б.Г. Экспертная информация. Методы получения и анализа. М.: Радио и связь, 1982. 184 с. 3. Мороз Б.И., Ульяновская Ю.В. Анализ мер близости объектов для различных типов экспертной информации // АСУ и приборы автоматики. 2008. Вып 144. С. 194-198.

Поступила в редколлегию 22.05.2011

Ульяновская Юлия Викторовна, доцент кафедры информационных систем и технологий Академии таможенной службы Украины. Научные интересы: экспертные системы, методы работы с нечеткими данными, экспертные оценки. Адрес: Украина, 49000, Днепрпетровск, ул. Роголёва, 8, тел.: (0562) 45-21-68.

МЕТОД ПРОГРАММНОГО РЕГУЛИРОВАНИЯ ИНЕРЦИОННЫМ ОБЪЕКТОМ ДЛЯ СИСТЕМЫ С ОДНИМ ДАТЧИКОМ И ОДНИМ ИСПОЛНИТЕЛЬНЫМ УСТРОЙСТВОМ

Разрабатывается математическая модель экстраполятора для системы программного управления, а также метод регулирования системы программного управления для тепловых объектов, которые имеют свойства линейного объекта с самовыравниванием. Метод имеет представление в виде рекуррентной формулы, где учитывается переходный процесс в объекте с помощью массива коэффициентов соответствия. Он разрабатывался для объектов с одним датчиком и одним исполнительным устройством.

Введение

Постановка задачи. Процесс управления есть упорядоченная логическая последовательность действий, связанная с переводом объекта управления из одного состояния в другое, призванная компенсировать накопленные на предыдущих этапах негативные тенденции и обеспечить достижение конечных результатов при заданном уровне экономической эффективности. Как и всякий технологический процесс, процесс управления не реализуется мгновенно, а имеет определенную протяженность во времени, зависящую от конкретных условий выполнения регулирования, поэтому возникает необходимость учитывать потери времени. Характерная черта адаптации - это текущее накопление информации о процессе функционирования системы, внешней среды и ее использование для улучшения избранного показателя качества. Процесс накопления информации связан с затратами времени, что в итоге приводит к запаздыванию в получении системой управления информации, необходимой для принятия решений. Это существенно снижает эффективность работы систем управления в реальном масштабе времени. Поэтому актуальной является задача прогнозирования состояний (ситуаций) системы, внешней среды и характеристик (поведения) системы для адаптивного управления. Такой прогноз может быть выполнен при использовании данного метода в системе управления. Для инерционных объектов часто применяют структуру, включающую идеальное звено экстраполяции. Теория управления с предсказанием рассматривает методы экстраполяции (предсказания) состояния объекта и выработки управляющего воздействия с упреждением [1, 2].

Анализ свойств объекта регулирования, требований к системе регулирования и характера входного сигнала системы $X(t)$ позволяет выделить следующие их особенности, существенно упрощающие синтез системы регулирования:

– объект регулирования должен обладать большой инерционностью, поэтому переход от непрерывной системы к дискретной не вызовет существенного ухудшения характеристик системы;

– объект управления должен относиться к классу объектов с самовыравниванием;

– теплофизические параметры объекта управления в заданном диапазоне температур должны оставаться неизменными;

– входной сигнал $X(t)$ – детерминированная наперед заданная функция времени;

– уровень помех на входе системы, к которым, очевидно, нужно отнести ошибку задания входной величины $X(t)$ и возмущающие воздействия (колебания температуры окружающей среды) на систему, небрежно мал;

– функция $X(t)$ не должна иметь производных по времени, равных бесконечности.

Это позволяет сделать следующие упрощения. Непрерывная величина $X(t)$, которая отражает закон изменения температуры в объекте, заменяется последовательностью дискретных значений X_1, X_2, \dots, X_m . Переходная функция объекта $h(t)$ заменяется набором дискретных значений K_1, K_2, \dots, K_n . Период дискретизации функций $X(t)$ и $h(t)$ одинаков и равен $\tau = \text{const}$. Эти упрощения позволили синтезировать структурную схему системы терморегулирования.

Цели и задачи: разработать метод программного регулирования с предсказанием для инерционного объекта управления с одним датчиком и одним исполнительным устройством при условии, что объекту присущи свойства линейности, самовыравнивания и справедлив принцип суперпозиции.

Анализ литературы. Истоки идеи индуктивного моделирования кроются в проблеме синтеза оптимального нелинейного предсказывающего фильтра, которую впервые сформулировал А. Н. Колмогоров в 1941г. [1]. Дальнейшее развитие идея получила в теории линейной фильтрации Колмогорова-Винера [1,2]. В начале 60-х годов прошлого века Д. Габор предложил универсальный предсказывающий фильтр с самонастройкой в процессе обучения [3], который реализует алгоритм предсказания будущего значения стационарной функции времени по ее предыстории путем нахождения оптимальных весовых коэффициентов расширенного оператора предсказания. Однако перечисленные работы не содержат моделей фильтров, предназначенных для решения задач управления тепловыми объектами. Некоторые вопросы теории предсказания детерминированных и случайных процессов рассмотрены в работе [5], где особое внимание уделяется реализации различных алгоритмов-операторов предсказания на электронных цифровых вычислительных машинах. Результаты этой работы также не доведены до формы, удобной для управления процессом нагрева/охлаждения инерционных объектов. В работе [6] предложено несколько методов управления температурными полями, однако в них не используется предсказывающий фильтр.

Математическая модель метода управления

Для реализации предлагаемого метода на стадии подготовки необходимо снять переходную характеристику объекта регулирования. По полученной кривой измеряется длительность переходных процессов в объекте (рис. 1). По определению, переходная функция объекта регулирования есть реакция объекта на управляющее воздействие в виде единичной функции [3]. Для линейных объектов справедливо утверждение, что характер переходной функции не зависит от величины (амплитуды) управляющего воздействия, т.е. отношение выходного параметра объекта $Y(t)$ (приращения температуры) к управляющему воздействию X вида единичной функции есть величина постоянная для одного и того же момента времени, для всех X :

$$K = \frac{Y(t)}{X} \Big|_{t=\text{const}} = \text{const} \quad (0 < X < X_{\text{max}}), \quad (1)$$

где X_{max} – максимальное значение управляющего воздействия, при котором сохраняются линейные свойства объекта управления. Очевидно, что с течением времени значение

коэффициента K будет изменяться вплоть до окончания переходного процесса в объекте. Если переходную функцию объекта по оси времени разбить на n равных интервалов τ (рис.1, $n=6$), то в моменты времени t , кратные τ , можно рассчитать коэффициенты K_j по формуле:

$$K_j = \frac{Y(t)}{X} \Big|_{t=\tau \cdot j} \quad (1 < j < n), \quad (2)$$

которыми однозначно определяется переходная функция объекта в этих точках для любых значений управляющего воздействия:

$$Y(t) = X \cdot K_j \Big|_{t=\tau \cdot j} \quad (0 < X < X_{\max}; 1 < j < n). \quad (3)$$

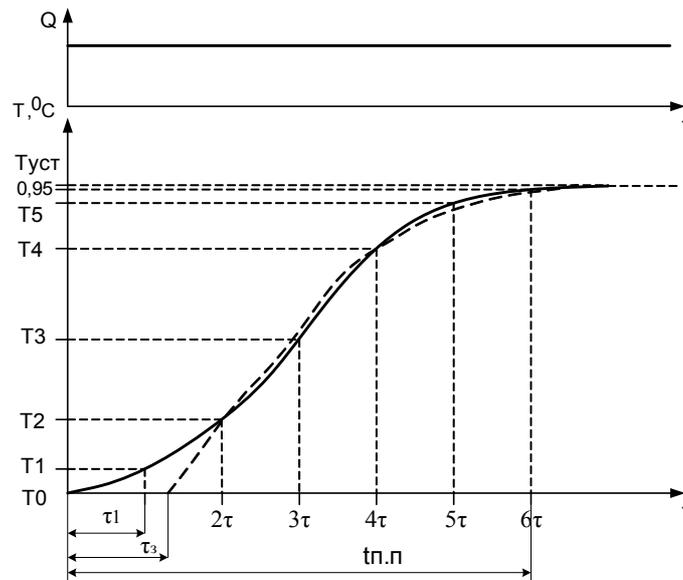


Рис. 1. Переходная функция исследуемого объекта

С другой стороны, с помощью коэффициентов K_j можно рассчитать значение управляющего воздействия X , которое за заданное время $t = i \cdot \tau$ вызовет изменение выходного параметра до $Y(i)$:

$$X = \frac{Y(i)}{K_j} \Big|_{i=\tau \cdot j} \quad (1 < j < n). \quad (4)$$

Таким образом, задание переходной характеристики в виде матрицы коэффициентов $[K]$ позволяет однозначно связать величину управляющего воздействия, заданного в виде единичной функции, с состоянием объекта управления.

В соответствии с принципом суперпозиции выходную величину можно рассматривать как алгебраическую сумму реакций объекта на элементарные управляющие воздействия, алгебраической суммой которых можно представить любое управляющее воздействие. На основании сказанного выше любую ступенчатую функцию, в виде которой сформировано управляющее воздействие, можно представить в виде суммы единичных функций:

$$X_i = x_0 + x_1 + x_2 + \dots + x_i \quad (0 < i < \infty), \quad (5)$$

а реакцию объекта вычислять как сумму реакций на соответствующие единичные функции:

$$Y_i = y_0 + y_1 + y_2 + \dots + y_i \quad (0 < i < \infty). \quad (6)$$

На рис. 2 представлена структурная схема системы, реализующая предлагаемый метод [4]. Данная система содержит блок 1 памяти кодов приращения тепловых потоков; блок 2 памяти кодов коэффициентов K_j ; вычислитель 3 прогнозируемого изменения температуры к концу интервала регулирования под воздействием тепла, подведенного к началу этого

интервала; программный задатчик 4; элемент сравнения 5 для вычисления кода и знака прогнозируемой ошибки рассогласования; аналого-цифровой преобразователь 6; элемент сравнения 7 для вычисления кода и знака имеющейся ошибки рассогласования; сумматор 8 для вычисления суммарной прогнозируемой ошибки рассогласования; вычислитель 9 для вычисления кодов приращения теплового потока и суммарного теплового потока. Система также содержит функциональный преобразователь 10 для вычисления кода, пропорционального току, который необходимо подать на нагреватель в течение следующего интервала программного регулирования; элемент памяти 11 для хранения кода, пропорционального току в нагревателе; цифро-аналоговые преобразователи 12, усилитель тока 13, нагреватель 14, объект регулирования 15 и датчик температуры 16.

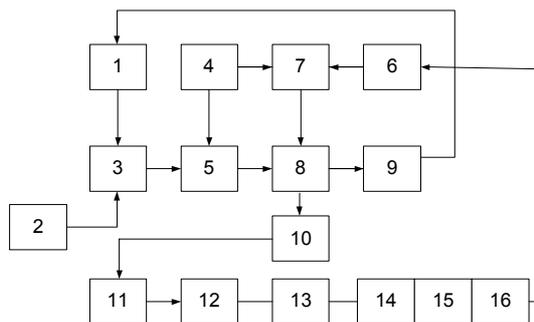


Рис. 2. Структурная схема системы программного регулирования объектом

В исходном состоянии блок памяти 1 и элемент памяти 11 обнулены, в блок памяти 2 занесены коды K_j . В блок задатчика 4 занесена программа изменения температуры объекта через интервалы времени τ в виде кодов приращения температуры относительно начальной температуры T_0 . На первый вход элемента сравнения 5 подается код приращения температуры, соответствующий концу первого интервала τ программного регулирования, на первый вход элемента сравнения 7 – нулевой код, т.е. код приращения температуры объекта в начальный момент времени. На второй вход элемента сравнения 7 подается код отклонения температуры объекта от T_0 . После запуска системы программного регулирования вычислитель начинает вычисление прогнозируемого изменения температуры объекта относительно T_0 . Для вычисления величины управляющего воздействия, которое за промежуток времени τ выведет объект в точку, заданную по программе, необходимо вычислить, в какую точку выйдет объект под управляющим воздействием, которое имело место до начала текущего интервала времени τ . Поэтому прогнозируемое изменение температуры в точке $t = (i+1) \cdot \tau$ необходимо вычислять без учета теплового потока, который будет подан в момент, когда время превысит $t \cdot \tau$. С учетом этого формула примет вид:

$$\Delta T_{i+1}^P = K_n \sum_{j=0}^{i-n+1} \Delta Q_j + \sum_{m=i-n+2}^{i-1} \Delta Q_m \cdot K_{i-m+1}, \quad i=1,2,\dots,\infty, \quad (7)$$

где ΔT_{i+1}^P – расчетное прогнозируемое изменение температуры объекта к моменту времени $t = (i+1) \cdot \tau$ под воздействием суммарного теплового потока, подведенного до момента времени $t = i \cdot \tau$; ΔQ_j – приращение управляющего воздействия теплового потока в момент времени $t = j \cdot \tau$. Формула представляет собой алгоритм работы вычислителя 3. Графическая интерпретация вычисления прогнозируемого изменения температуры приведена на (рис. 3). Для наглядности иллюстрации число n ограничено шестью. Управляющее воздействие (график 1) можно разбить на единичные функции (графики 2, 4, 6, 8). На графиках 3, 5, 7, 9 представлены изменения температуры объекта под воздействием соответствующих тепловых потоков. Суммарное изменение температуры объекта представлено на графике 12 и вычисляется как алгебраическая сумма ординат графиков 3, 5, 7, 9 [5].

Рассчитаем прогнозируемое изменение температуры объекта в момент времени $t = 10\tau$:

$$\Delta T_{10}^P = (\Delta Q_0 + \Delta Q_1 + \Delta Q_2 + \Delta Q_3 + \Delta Q_4)K_6 + \Delta Q_5 \cdot K_5 + \Delta Q_6 \cdot K_4 + \Delta Q_7 \cdot K_3 + \Delta Q_8 \cdot K_2. \quad (8)$$

В приведенном на рис. 3 примере ΔQ в моменты времени $3\tau, 5\tau, 7\tau, 8\tau, 9\tau$ равны нулю.

В элементе сравнения 5 код расчетной температуры сравнивается с кодом температуры датчика 4. Код разности с соответствующим знаком, равный:

$$\Delta_1 = \Delta T_{i+1}^3 - \Delta T_{i+1}^P, \quad (9)$$

где ΔT_{i+1}^3 – приращение температуры, требуемое по программе, поступает на первый вход сумматора 8.

Во время выполнения программы регулирования температура окружающей среды может изменяться. Так как скорость изменения температуры окружающей среды обычно значительно меньше скорости температуры объекта, переходными процессами, обусловленными колебаниями внешней температуры, можно пренебречь. Однако эти колебания могут вызвать ошибку рассогласования. Ошибки округления при кодировании K_j и ΔQ_i , равные:

$$\delta = \frac{A}{2^n}, \quad (10)$$

где A – максимальное значение кодируемой величины, 2^n – число разрядов регистров памяти, а также погрешности ЦАП и АЦП могут вызвать ошибку рассогласования, которую необходимо учитывать. С этой целью на элемент сравнения 7 подается код приращения температуры датчика для момента времени $t = i \cdot \tau$ и код температуры датчика, который поступает с выхода АЦП 6. Разностный код, равный:

$$\Delta_2 = \Delta T_i^3 - \Delta T_i^g, \quad (11)$$

подается на второй вход сумматора 8. Так как изменение внешней температуры предсказать сложно и поскольку это изменение за время τ пренебрежимо мало, можно принять $\Delta_2(i+1) = \Delta_2(i)$. Исходя из этого рассчитывается полное отклонение прогнозируемой температуры объекта от температуры датчика:

$$\Delta = \Delta_1 + \Delta_2 = \left(\Delta T_{i+1}^3 - \Delta T_{i+1}^P \right) + \left(\Delta T_i^3 - \Delta T_i^g \right). \quad (12)$$

Значение Δ вычисляется сумматором 8. Очевидно, что для $i=1$ при условии, что температура объекта не отличалась от T_0 , $\Delta T = \Delta T_{i+1}^3$, поскольку все остальные слагаемые равны нулю.

Для того чтобы с момента времени $t = (i+1) \cdot \tau$ вывести объект в заданную точку, начиная с момента времени $t = i \cdot \tau$ к объекту нужно подвести дополнительно такой тепловой поток, который вызовет равное по величине, но противоположное по знаку изменение температуры.

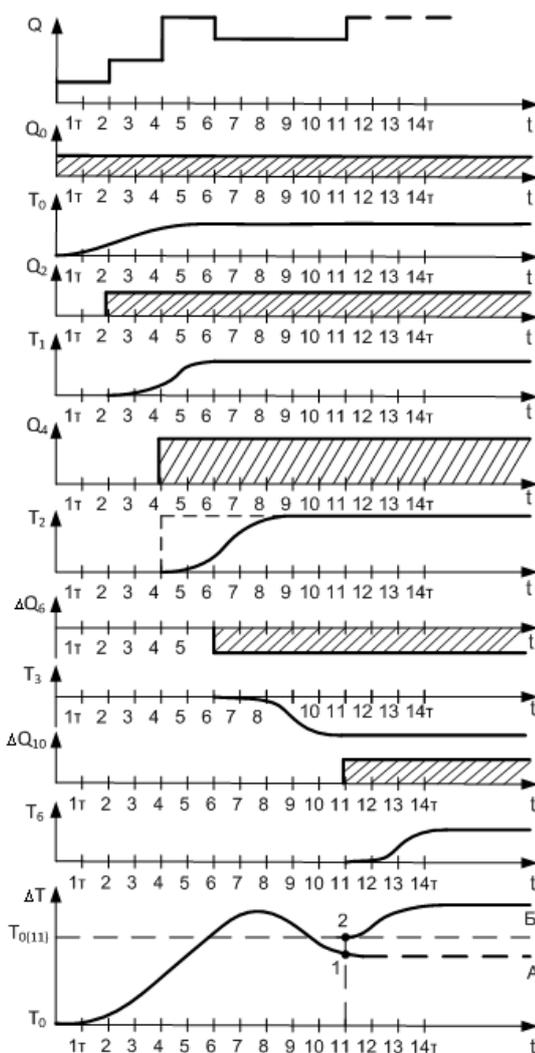


Рис. 3. Графическая интерпретация вычисления прогнозируемого изменения температуры

Значение приращения теплового потока вычисляется по формуле:

$$\Delta Q_i = \frac{\Delta}{K_1}, \quad (13)$$

где ΔQ_i – приращение теплового потока в момент времени $t = i \cdot \tau$. Поскольку полное отклонение температуры Δ необходимо компенсировать за время, равное τ , Δ необходимо разделить на K_1 .

Полный тепловой поток, который будет подаваться с момента времени $t = i \cdot \tau$ до $t = (i + 1) \cdot \tau$, равен:

$$Q_i = \sum_{j=1}^i \Delta Q_j. \quad (14)$$

Значение кодов приращения теплового потока ΔQ_i и суммарного теплового потока Q_i вычисляется в вычислителе 9, после чего код приращения пересылается в блок памяти 1, а код суммарного теплового потока поступает на вход функционального преобразователя 10. С выхода функционального преобразователя код, пропорциональный току нагревателя, записывается в элемент памяти 11, а затем преобразуется цифро-аналоговым преобразователем 12 в аналоговый сигнал, который усиливается усилителем мощности 13 и подается на нагреватель 14.

Под воздействием подведенного к объекту тепла объект начинает изменять свою температуру до ΔT_{i+1}^3 . После записи кода в элемент памяти 11 система начинает вычислять значение управляющего воздействия для интервала времени от $t = (i + 1) \cdot \tau$ до $t = (i + 2) \cdot \tau$.

Исследование реакции объекта на управляющие воздействия ступенчатой формы, расчет коэффициентов K , экспериментальное подтверждение справедливости полученных выражений произведены с применением тепловой модели в среде Ansys [5, 6].

Выводы

Научная новизна. Приведены формулы, описывающие процесс управления инерционным объектом. Выведены формулы, описывающие программное регулирование с предсказанием инерционным объектом для системы с одним датчиком и одним исполнительным устройством при подаче на нее управляющего воздействия ступенчатой формы.

Практическая ценность работы заключается в том, что рассмотренная математическая модель достаточно точно описывает процесс управления с предсказанием. На основе математической модели разработана структурная схема и алгоритм функционирования системы управления с предсказывающим фильтром.

Перспективы исследований. Планируется провести исследования и применить данный подход управления с предсказанием поведения объекта для управления системами с распределенными параметрами.

Список литературы: 1. Колмогоров А. Н. Проблема синтеза оптимального предсказывающего фильтра. Изв. АН СССР. Сер. матем. и естеств. наук. 1941. №5. С 112-129. 2. Weiner N. The Extrapolation Interpolation and Smoothing of Stationary Time-Series. I. Wiley, N.Y., 1949. 290 p. 3. Gabor D., Wilby W.R., Woodcock R.A. A universal nonlinear filter, predictor and simulator which optimizes itself by a learning process. Proc. Inst. Electr. Engrs. 1961. vol. 108. part B, №40, P. 85-98. 4. Ивахненко А. Г., Лапа В. Г. Предсказание случайных процессов. Киев: Наук. думка, 1971. 415 с. 5. Сироджа И. Б. Квантовые модели и методы искусственного интеллекта для принятия решений и управления. Киев: Наук. думка, 2002. 490 с. 6. Бутковский А. Г. Методы управления системами с распределенными параметрами. М.: Наука, 1975. 568 с.

Поступила в редколлегию 11.06.2011

Гапон Анатолий Иванович, декан факультета АП НТУ «ХПИ», канд. техн. наук., доцент кафедры АУТС НТУ «ХПИ». Научные интересы: теория управления, системы управления объектами с распределенными параметрами. Увлечения: баскетбол, путешествия. Адрес: Украина, 61002, Харьков, ул. Фрунзе, 21, тел. (057) 707-62-66. E-mail: gapon@kpi.khtrpov.ua

Савицкий Сергей Михайлович, аспирант кафедры АУТС НТУ «ХПИ». Научные интересы: теория управления, системы управления объектами с распределенными параметрами, инерционные объекты, математическое моделирование. Увлечение: футбол, плавание, баскетбол, шахматы, русский бильярд, горные лыжи, общественная деятельность. Адрес: Украина, 61002, Харьков, ул. Фрунзе, 21, тел.: (057) 707-62-66. E-mail: savitskiy.s@ukr.net