ИНИСТЕРСТВО ЦИФРОВОГО РАЗВИТИЯ, СВЯЗИ И МАССОВЫХ КОММУНИКАЦИЙ РОССИЙСКОЙ ФЕДЕРАЦИИ ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «СИБИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ТЕЛЕКОММУНИКАЦИЙ И ИНФОРМАТИКИ»

Вестник СибГУТИ Том 16, № 2, 2022

Выпускается ежеквартально, выходит с 2007 г.

Учредитель -

федеральное государственное бюджетное образовательное учреждение высшего образования «Сибирский государственный университет телекоммуникаций и информатики»

В. И. Носов, д.т.н., проф. С. В. Поршнев, д.т.н., проф.

А. С. Родионов, д.т.н., доц. А. И. Романенко, д.ф.-м.н., проф.

И. И. Рябцев, д.ф.-м.н., чл.-корр. РАН

Б. Я. Рябко, д.т.н., проф.

О. В. Стукач, д.т.н., доц.

В. К. Трофимов, д.т.н., проф.

С. В. Федоренко, д.т.н., доц.

Ю. И. Шокин, д.ф.-м.н., акад. РАН

А. И. Фалько, д.т.н., проф.

Б. Г. Хаиров, д.э.н., доц. А. Г. Черевко, к.ф.-м.н., доц.

С. В. Шидловский, д.т.н.

В. П. Шувалов, д.т.н., проф.

Э. Сименс, д.т.н.

Председатель редакционного совета А. Н. Фионов, д.т.н., проф.

Редакционный совет:

- С. С. Абрамов, д.т.н., доц.
- В. Б. Барахнин, д.т.н., доц.
- В. М. Белов, д.т.н., проф.
- В. Н. Васюков, д.т.н., проф.
- В. Ю. Васильев, д.х.н., доц.
- Н. И. Горлов, д.т.н., проф.
- Н. Л. Казначеева, д.э.н., доц.
- В. С. Канев, д.т.н., проф.
- А. И. Карпович, д.э.н., проф.
- М. Г. Курносов, д.т.н., проф.
- В. В. Лебедянцев, д.т.н., проф.
- А. В. Лихачёв, д.т.н.
- С. Н. Мамойленко, д.т.н., доц.
- О. Г. Мелентьев, д.т.н., проф.
- Р. В. Мещеряков, д.т.н., проф.
- И. Г. Неизвестный, д.ф.-м.н., чл.-корр. РАН
- С. Н. Новиков, д.т.н., доц.

Редакция:

А. Н. Фионов (главный редактор), М. Ю. Галкина (заведующая редакцией),

Н. А. Двуреченская (технический и литературный редактор, компьютерная вёрстка),

Т. А. Алфёрова (лингвист-корректор).

Адрес редакции и издателя

630102, г. Новосибирск, ул. Кирова, д. 86

e-mail: vestnik@sibguti.ru

Информация о журнале доступна в сети Internet по адресу https://sibsutis.elpub.ru/jour.

Журнал включён в Перечень ВАК российских рецензируемых научных журналов, в которых должны быть опубликованы основные научные результаты диссертаций на соискание учёных степеней доктора и кандидата наук.

Журнал зарегистрирован Федеральной службой по надзору за соблюдением законодательства в сфере массовых коммуникаций и охране культурного наследия, свидетельство о регистрации ПИ № ФС77-25835 от 29.09.2006. ISSN 1998-6920. Подписной индекс в объединённом каталоге «Пресса России» – 82519.

Дата выхода в свет 21.07.2022. Отпускная (редакционная) цена 800 руб.

Отпечатано в издательском центре СибГУТИ по адресу: 630102, г. Новосибирск, ул. Кирова, д. 86. Бумага офсетная, формат А4. Тираж 300 экз.

© СибГУТИ, 2022 г.

СОДЕРЖАНИЕ

У. В. Павлова, А. А. Ракитский Применение специализированных детерминированных конечных автоматов для протнозирования временных рядов 1. В. О. Доценко, И. Е. Шевнина 1. Исследование актуальных требований для разработки систем мониторинга 2. и управления телекоммуникационной сети 2. И. Д. Иогансон, А. А. Голованов, ЖМ. Н. Дакуо, В. В. Давыдов 2. Криптографический протокол поиска места встречи участников со свойством 3. В. И. Носов, А. С. Ладан, М. В. Зиновьев 3. Определение оптимальной длины кадра в сети наземного цифрового 4. С. Б. Жанаева 5. Сбор данных о работе оборудования в сети мобильного оператора связи 5. А. Н. Сычук, В. А. Варданян 9. Оценка помехоустойчивости DWDM-сигналов в нелинейном режиме 4. Функционирования волоконно-оптического тракта 6. В. Г. Арсентьев, Г. И. Криволапов 7. Повышение точности позиционирования в навигационной системе 7. Алгоритм сбалансированной интерполяции пикселей при выделении 7. В. В. Ерохин 8. 8. Верификация модели интегральной катушки индуктивности 7. В. В. Ерохин 5. <	М. Г. Курносов Иерархический алгоритм барьерной синхронизации для многопроцессорных систем с общей памятью	3
В. О. Доценко, И. Е. Шевнина Исследование актуальных требований для разработки систем мониторинта и управления телекоммуникационной сети 2. И. Д. Иогансон, А. А. Голованов, ЖМ. Н. Дакуо, В. В. Давыдов 8. Криптографический протокол поиска места встречи участников со свойством 3. В. И. Носов, А. С. Ладан, М. В. Зиновьев 3. Определение оптимальной длины кадра в сети наземного цифрового 4. С. Б. Жанаева 4. Сбор данных о работе оборудования в сети мобильного оператора связи 5. А. Н. Сычук, В. А. Варданян 5. Оценка помехоустойчивости DWDM-сигналов в нелинейном режиме 6. В. Г. Арсентьев, Г. И. Криволапов 6. Повышение точности позиционирования в навигационной системе 6. В. Г. Арсентьев, Г. И. Криволапов 7. Повышение точности позиционирования в навигационной системе 7. В. В. Заерко, Н. Л. Боброва 7. Алгоритм сбалансированной интерполяции пикселей при выделении контуров объектов на границах изображения 7. В. В. Ерохин 7. Верификация модели интегральной катушки индуктивности для СВЧ LC-фильтров в Si- и SiGe-системах на кристалле 9.	У. В. Павлова, А. А. Ракитский Применение специализированных детерминированных конечных автоматов для прогнозирования временных рядов	12
И. Д. Иогансон, А. А. Голованов, ЖМ. Н. Дакуо, В. В. Давыдов Криптографический протокол поиска места встречи участников со свойством Конфиденциальности 3. В. И. Носов, А. С. Ладан, М. В. Зиновьев 9 Определение оптимальной длины кадра в сети наземного цифрового 4 С. Б. Жанаева 4 Сбор данных о работе оборудования в сети мобильного оператора связи 5 А. Н. Сычук, В. А. Варданян 5 Оценка помехоустойчивости DWDM-сигналов в нелинейном режиме 6 В. Г. Арсентьев, Г. И. Криволапов 6 Повышение точности позиционирования в навигационной системе 7 А. В. Заерко, Н. Л. Боброва 7 Алгоритм сбалансированной интерполяции пикселей при выделении контуров объектов на границах изображения 7 В. В. Ерохин 7 В. В. Ерохин 9 Верификация модели интегральной катушки индуктивности 9	В. О. Доценко, И. Е. Шевнина Исследование актуальных требований для разработки систем мониторинга и управления телекоммуникационной сети	23
В. И. Носов, А. С. Ладан, М. В. Зиновьев Определение оптимальной длины кадра в сети наземного цифрового телевизионного вещания 4 С. Б. Жанаева 4 Сбор данных о работе оборудования в сети мобильного оператора связи 5 А. Н. Сычук, В. А. Варданян 5 Оценка помехоустойчивости DWDM-сигналов в нелинейном режиме 6 В. Г. Арсентьев, Г. И. Криволапов 6 Повышение точности позиционирования в навигационной системе 7 Д. В. Заерко, Н. Л. Боброва 7 Алгоритм сбалансированной интерполяции пикселей при выделении контуров объектов на границах изображения 7 В. В. Ерохин 7 В. В. Ерохин 8 6 В. С. С. Ч. Ц. Сфильтров в Si- и SiGe-системах на кристалле 9	И. Д. Иогансон, А. А. Голованов, ЖМ. Н. Дакуо, В. В. Давыдов Криптографический протокол поиска места встречи участников со свойством конфиденциальности	33
 С. Б. Жанаева Сбор данных о работе оборудования в сети мобильного оператора связи	В. И. Носов, А. С. Ладан, М. В. Зиновьев Определение оптимальной длины кадра в сети наземного цифрового телевизионного вещания	40
А. Н. Сычук, В. А. Варданян Оценка помехоустойчивости DWDM-сигналов в нелинейном режиме функционирования волоконно-оптического тракта	С. Б. Жанаева Сбор данных о работе оборудования в сети мобильного оператора связи	55
 В. Г. Арсентьев, Г. И. Криволапов Повышение точности позиционирования в навигационной системе автономного необитаемого подводного аппарата	А. Н. Сычук, В. А. Варданян Оценка помехоустойчивости DWDM-сигналов в нелинейном режиме функционирования волоконно-оптического тракта	63
 Д. В. Заерко, Н. Л. Боброва Алгоритм сбалансированной интерполяции пикселей при выделении контуров объектов на границах изображения	В. Г. Арсентьев, Г. И. Криволапов Повышение точности позиционирования в навигационной системе автономного необитаемого подводного аппарата	72
В. В. Ерохин Верификация модели интегральной катушки индуктивности для СВЧ LC-фильтров в Si- и SiGe-системах на кристалле	Д. В. Заерко, Н. Л. Боброва Алгоритм сбалансированной интерполяции пикселей при выделении контуров объектов на границах изображения	79
	В. В. Ерохин Верификация модели интегральной катушки индуктивности для СВЧ LC-фильтров в Si- и SiGe-системах на кристалле	94

Иерархический алгоритм барьерной синхронизации для многопроцессорных систем с общей памятью

М. Г. Курносов¹

Разработан иерархический алгоритм барьерной синхронизации стандарта MPI, который формирует группы процессов, разделяющие общие ресурсы на уровнях иерархии памяти: кеш-память L2/L3, NUMA-узел, процессор. Синхронизация выполняется параллельно в группах на каждом уровне иерархии, что позволяет локализовать межпроцессные взаимодействия. Реализация выполнена на базе библиотеки Open MPI. Эксперименты на сервере с двумя процессорами Huawei Kunpeng (128 ядер, 4 NUMA-узла) показали, что предложенный алгоритм с группировкой процессов по NUMA-узлам обеспечивает минимальное время выполнения по сравнению с известными методами и устойчив к изменению способа распределения процессов по процессорным ядрам.

Ключевые слова: барьер, синхронизация, МРІ.

1. Введение

В работе рассматриваются методы реализации операции барьерной синхронизации стандарта MPI в пределах одного многопроцессорного сервера (вычислительного узла). Такие алгоритмы применяются, если MPI-программа запущена на одном узле или в составе иерархических (многоуровневых) алгоритмов коллективных операций стандарта MPI [1]. Большая часть современных многопроцессорных серверов имеет архитектуру с неоднородным доступом к памяти. В таких системах множество процессорных ядер разбито на NUMA-узлы, каждый из которых имеет свой интегрированный контроллер доступа к локальной оперативной памяти. Доступ к памяти удаленных NUMA-узлов осуществляется через межпроцессорный канал связи (например, Intel UPI, AMD InfinityFabric, HiSilicon Hydra) и требует больше времени. Количество NUMA-узлов на многопроцессорном сервере зависит от микроархитектуры процессора, числа процессоров и активированного в BIOS режима разделения системы на NUMA-узлы (Intel Sub-NUMA Clustering, AMD EPYC Channel-Interleaving, Huawei Channel Interleaving). На рис. 1 показана система с четырьмя NUMA-узлами – два процессора Huawei Kunpeng 920-6426 (64 ядра с архитектурой ARMv8, 2 NUMA-узла на процессор). Обращение к страницам памяти удаленного процессора требует перехода по шине HiSilicon Hydra.

¹ Работа выполнена в рамках Государственного задания № 071-03-2022-001.



Рис. 1. NUMA-архитектура сервера на базе двух процессоров Huawei Kunpeng 920-6426 (справа показана матрица расстояний между NUMA-узлами: numactl -H)

Операция барьерной синхронизации MPI_Barrier блокирует выполнение процесса MPIпрограммы до тех пор, пока каждый процесс коммуникатора не осуществит её вызов. Процесс выходит из барьерной синхронизации только после того, как все процессы коммуникатора начали её выполнение. Известные алгоритмы барьерной синхронизации для многопроцессорных систем с общей памятью: глобальный счетчик (central counter) [1–2], плоское дерево [1], плоское дерево с фазами gather/release [1], объединяющее дерево (combining tree) [3], MCSбарьер [4], турнирный алгоритм (tournament) [5], рассеивающий (dissemination) [6] основаны на использовании счетчиков и флагов в сегменте разделяемой памяти, при помощи которых процессы уведомляют друг друга о достижении барьера или его определенного этапа. Время выполнения барьера зависит от времени доступа к разделяемым переменным, а это, в свою очередь, зависит и от распределения процессов по ядрам системы.

В данной работе предложен алгоритм, который группирует процессы, совместно использующие ресурсы на каждом уровне иерархии памяти (разделяют кеш-память L2, L3, контроллер доступа к памяти NUMA-узла, принадлежат одному процессору). Это позволяет локализовать синхронизацию в созданных группах и тем самым сократить время операции MPI_Barrier.

2. Иерархический алгоритм барьерной синхронизации

2.1. Построение иерархии групп процессов

При создании нового МРІ-коммуникатора процессы формируют совместно используемый сегмент разделяемой памяти (вызовом mmap () операционной системы GNU/Linux), который затем используется для взаимодействия процессов в ходе выполнения алгоритмов. Часть сегмента отведена под блок – непрерывную область в виртуальном адресном пространстве с глобальными счетчиками и флагами, а также под индивидуальные блоки процессов. Длины блоков всех процессов одинаковы и кратны размеру страницы физической памяти. Зная свой номер rank и размер блока, каждый процесс вычисляет начало своего участка в разделяемом сегменте. В блоке каждого процесса хранятся его локальные счетчики и флаги. Чтобы избежать ложного разделения данных (false sharing), под каждый счетчик/флаг отводится область, длина которой равна размеру строки кеш-памяти целевой архитектуры (для Intel 64 – 64 байта, для Kunpeng 920-6426 ARMv8 – 128 байт), с соответствующим выравниванием начального адреса. Для сокращения времени доступа к данным своего блока каждый процесс пытается выделить физические страницы памяти с локального NUMA-узла – каждый процесс самостоятельно инициализирует счетчики в своем блоке. В соответствии с политикой first touch policy ядра Linux это приводит к выделению страницы физической памяти с NUMA-узла ядра, на котором выполняется процесс, осуществивший первую запись в страницу памяти. По умолчанию блок с глобальными счетчиками инициализирует процесс 0 (глобальные счетчики и флаги будут размещены в памяти его NUMA-узла).

После создания сегмента памяти каждый процесс осуществляет анализ распределения процессов по ядрам и построение групп процессов, соответствующих иерархии памяти вычислительного узла.

Шаг 1. Анализ иерархии памяти. МРІ-процесс определяет число ядер, разделяющих кешпамять уровней L2, L3, число NUMA-узлов и процессоров на вычислительном узле. Вычисляется количество *L* и формируется список активных уровней иерархии памяти – уровней, которые совместно используются двумя и более ядрами: кеш-память L2, L3, NUMA-узел, процессор (socket, package). В текущей реализации информация получается из библиотеки HWLOC. На рис. 2 приведен пример двухпроцессорного сервера, в котором кеш-память L1 и L2 у каждого ядра своя, поэтому список активных уровней составляют NUMA-узлы (32 ядра) и процессоры (PACKAGE, 2 NUMA-узла, 64 ядра). Уровень кеш-памяти L3 не учтён, так как те же ядра разделяют один NUMA-узел.



Рис. 2. Двухпроцессорный сервер: 2 NUMA-узла на процессор, 32 ядра на NUMA-узел; 128 MPI-процессов с привязкой к процессорным ядрам (--bind-to core) и последовательным распределением (--map-by core), два активных уровня иерархии (*L* = 2), три уровня групп процессов

Шаг 2. Определение таблицы размещения процессов в пределах узла. Процесс определяет множество процессорных ядер, к которым он привязан (на каких ядрах планировщик операционной системы GNU/Linux может его выполнять). По умолчанию библиотека Open MPI привязывает процесс к процессорному ядру (--bind-to core), однако при запуске программы пользователь может изменить способ привязки – разрешить операционной системе динамически переносить процесс в пределах одного процессора (--bind-to package), в пределах NUMA-узла (--bind-to numa) или отключить привязку (--bind-to none) и разрешить выполнение процесса на любом ядре. При отключенной привязке процессов к процессорным ядрам алгоритм завершает свою работу, так как невозможно однозначно определить, какие уровни иерархии памяти процессы используют совместно. Информация о привязке процесса извлекается через интерфейс PMIx библиотеки Open MPI (PMIX LOCALITY STRING). Через обращение к интерфейсу PMIx определяется размещение всех *P* процессов коммуникатора в пределах узла. Для активных уровней иерархии памяти определяется номер функционального модуля procloc[l][i], который используется каждым процессом i на уровне l. Для примера, на рис. 2 PMIx-строка с размещением процесса 32 имеет вид "SK0:L31:L232:L132:CR32:HT32:NM1" (номер процессора: номер кеш-памяти L3:...: номер NUMA-узла). Процесс входит в две группы, таким образом, для процесса 32 procloc[0][32] = 1, procloc[1][32] = 0. Вычислительная сложность шага – $O(P \cdot L)$.

Шаг 3. Формирование групп процессов. Каждый процесс для всех активных уровней иерархии памяти строит списки процессов, с которыми он разделяет один функциональный модуль (кеш-память L2, NUMA-узел, процессор). Построение начинается с низшего активного уровня. В каждой группе выбирается лидер – процесс с наименьшим номером. Лидеры групп

становятся членами групп следующего активного уровня. Цикл повторяется для всех активных уровней. Вычислительная сложность шага – O(L(C + P)), где L – число активных уровней в иерархии, C – число ядер на вычислительном узле, P – число процессов в коммуникаторе. На рис. 2 показана иерархия из трех групп G1, G2, G3 для двух активных уровней (NUMA и PACKAGE). Группа G3 включает два процесса – лидеры процессоров 0 и 1. Ниже приведены примеры групп процессов:

- процесс 0: G1(0, 1, ..., 31), G2(0, 32), G3(0, 64);
- процесс 1: G1(0, 1, ..., 31);
- процесс 32: *G*1 (32, 33, ..., 63), *G*2(0, 32).

На рис. 3 показан пример иерархии групп для случая 14 процессов с циклическим распределением по 4 процесса на NUMA-узел.



Рис. 3. Двухпроцессорный сервер: 2 NUMA-узла на процессор, 32 ядра на NUMA-узел; 14 процессов с привязкой к процессорным ядрам (--bind-to core) и циклическим распределением процессов по NUMA-узлам (--map-by numa), два активных уровня иерархии (L = 2), три уровня групп процессов

2.2. Алгоритм барьерной синхронизации

В момент вызова операции MPI_Barrier каждому процессу известен свой исходный номер rank в коммуникаторе, а также номер group[l].group_rank в каждой его группе уровня l. В сегменте разделяемой памяти процесс хранит счетчики group[l].state[group_rank] числа обращений к барьеру на уровне иерархии $l \in \{0, 1, ..., L\}$, а также имеет доступ к общему флагу sense уведомления о достижении барьера каждым процессом. При входе в барьер каждый процесс rank меняет значение своего флага sense_local[rank] на противоположное. Далее выполняется синхронизация процессов в группах уровня 0, 1, ..., L. Синхронизация в группе выполняется алгоритмом плоского дерева – лидер ожидает, пока члены группы не переведут счетчики состояния в необходимое значение. После этого лидер переходит к синхронизации на уровне выше (BarrierGroupLevel (level + 1)), а остальные процессы переходят в ожидание, пока значение локального флага sense_local[rank] не станет равным значению глобального флага sense. Инвертирование счетчиков sense_local[rank] и sense обеспечивает корректную работу многократных вызовов барьера (схема sense reversing [2, 4]). Начальные состояния: sense = 1, sense local[rank] = 1. Ниже приведен псевдокод алгоритма.

```
algorithm MPI_Barrier()
sense_local[rank] = sense_local[rank] xor 1
BarrierGroupLevel(0)
end algorithm
algorithm BarrierGroupLevel(level)
grank = group[level].group_rank // Индекс процесса в группе
gleader = group[level].group_rank_leader // Индекс лидера в группе
group[level].state[grank]++ // Уведомление лидера о входе
```

```
if grank = gleader then
    while true do
                        // Лидер ожидает уведомления от процессов группы
     narrived = 0
      for child = 0 to group[level].size - 1 do
        if group[level].state[child] >= group[level].state[gleader] then
          narrived++
      end for
      if narrived = group[level].size then break
    end while
    if level + 1 < ngroups then</pre>
                                     // Лидер переходит на следующий уровень
     BarrierGroupLevel(level + 1)
    else if rank = 0 then
      sense = sense local[rank]
                                     // Лидер уведомляет о выходе из барьера
    end if
  else
    while sense != sense local[rank] do // Ожидание уведомления от лидера 0
  end if
end algorithm
```

Время выполнения алгоритма $O(L \cdot P)$ линейно зависит от числа процессов и активных уровней иерархии памяти. В программном компоненте реализована возможность отключения учета отдельных уровней иерархии памяти при формировании групп процессов. Например, алгоритм *shm-topo-NUMA-SOCKET* для системы на рис. 2 учитывает все активные уровни иерархии памяти. Алгоритм *shm-topo-NUMA* получен путем отключения учета уровня процессора, он группирует процессы только по NUMA-узлам (32 ядра в группе). Алгоритм *shm-topo-SOCKET* игнорирует NUMA-узлы и группирует 64 ядра каждого процессора в отдельную группу.

3. Результаты экспериментов

Эксперименты выполнялись на двухпроцессорном сервере: два процессора Huawei Kunpeng 920-6426 ARMv8 (64 ядра, 2 NUMA-узла на процессор, размер строки кеш-памяти L1/L2 – 64 байта, L3 – 128 байт), 255 Гб оперативной памяти, по 64 Гб на каждый NUMA-узел. Процессоры соединены каналами HiSilicon Hydra. На сервере установлена операционная система CentOS 7.6.1810 (ядро linux 4.14.0-115, aarch64, размер страницы физической памяти – 64 Кб), gcc 4.8.5. Матрица расстояний между NUMA-узлами показана на рис. 2.

Для реализации и тестирования алгоритмов использовалась библиотека Open MPI 5.0.0rc2 (параметры сборки: CFLAGS="-03 -g" --disable-debug). В качестве теста взят пакет OSU Micro-Benchmarks 5.8. Для каждого числа процессов операция MPI_Barrier запускалась 1000 раз (параметры запуска теста: osu_barrier -x1 -i1000 -f). В каждом эксперименте тест запускался 5 раз, отбрасывались наименьшее и наибольшее значения максимальной латентности, среди оставшихся значений отыскивалось среднее время работы операции [1, 2]. При запуске теста каждый процесс привязывался к процессорному ядру (mpiexec --bind-to core).

Выполнялось сравнение разработанного алгоритма операции MPI_Barrier с алгоритмами на базе операций send/recv библиотеки Open MPI (coll/base): base-liner (линейный алгоритм, время выполнения O(P)), base-double-ring (двухэтапный кольцевой алгоритм, O(P)), base-recursive-doubling (алгоритм рекурсивного удваивания, время $O(\log P)$), base-bruck (алгоритм Брука, $O(\log P)$), base-tree (двухэтапный алгоритм рекурсивного удваивания, $O(\log P)$). Также сравнение выполнялось с алгоритмами барьерной синхронизации, основанными на взаимодействии процессов через разделяемую память (shm-barrier) [2]: shm-sense-reversing (алгоритм с глобальным счетчиком на базе атомарных операций, время O(P)), shm-sense-reversingcounter (алгоритм плоского дерева, O(P)), shm-sense-reversing-gr (двухэтапный алгоритм плоского дерева, O(P)) [1], shm-combining-tree (алгоритм бинарного объединяющего дерева, O(logP)), shm-mcs (MCS-барьер), shm-tournament (турнирный алгоритм, O(logP)), shm-dissemination (рассеивающий алгоритм, O(logP)).

На рис. 4 показано время выполнения алгоритмов барьерной синхронизации на одном процессоре – 64 процесса в пределах одного NUMA-узла 0. Наименьшим временем выполнения характеризуется алгоритм бинарного объединяющего дерева (*shm-combining-tree*), который на 40 % быстрее турнирного алгоритма *shm-tournament* и на 55 % быстрее MCS-барьера. Производительность многоуровневого алгоритма на одном процессоре не исследовалась, так как в этом случае ядра разделяют единственный уровень иерархии – контроллер доступа к памяти NUMA-узла. Алгоритмы барьерной синхронизации на основе явного использования операций send/recv (*base*) значительно уступают по производительности алгоритмам, использующим для взаимодействий сегмент разделяемой памяти (*shm*). Алгоритм рекурсивного удваивания *base-tree* почти в 2.8 раз медленнее алгоритма объединяющего дерева. Исключение составляет алгоритм с *глобальным счетчиком* на базе атомарных операций (*shm-sense-reversing*), который демонстрирует низкую масштабируемость.



Рис. 4. Время выполнения операции MPI Ваггіег на 64 ядрах одного процессора

На рис. 5 приведены результаты для двух процессоров (128 ядер, четыре NUMA-узла). Наилучшие результаты демонстрирует иерархический алгоритм с группировкой процессов по NUMA-узлам (*shm-topo-NUMA*), незначительно ему уступает алгоритм объединяющего дерева. Иерархический алгоритм с группировкой по NUMA-узлам и процессорам (*shm-topo-NUMA-SOCKET*) опережает на 14 % алгоритм с группировкой только по сокетам (*shm-topo-SOCKET*). Алгоритм рекурсивного удваивания на базе операций send/recv *base-tree* в 2.5 раз медленнее иерархического алгоритма *shm-topo-NUMA*.



Рис. 5. Время выполнения операции MPI_Barrier на 128 ядрах двух процессоров (четыре NUMA-узла по 32 ядра)

До момента запуска MPI-приложения неизвестно, процессы с какими номерами на каких ядрах будут выполняться. На практике пользователи указывают одну из трех схем распределения процессов по ядрам:

- map-by core: процессы программы назначаются последовательно с шагом через одно ядро (процесс 0 на ядро 0, процесс 1 на ядро 1, ..., процесс 127 на ядро 127);
- тар-by NUMA: процессы циклически распределяются с шагом через NUMA-узел (для рис. 1 процесс 0 на ядро 0, процесс 1 на ядро 32, процесс 2 на ядро 64, процесс 3 на ядро 96, ..., процесс 127 на ядро 127);
- map-by SOCKET: процессы циклически распределяются по ядрам двух процессоров (шаг 64 ядра, на рис. 1 – процесс 0 на ядро 0, процесс 1 на ядро 64, процесс 2 на ядро 1, процесс 3 на ядро 65, ..., процесс 127 на ядро 127).

Способ распределения процессов может значительно влиять на время выполнения одного и того же алгоритма, так как в процессе синхронизации процессы могут читать и записывать флаги и счетчики, находящиеся в памяти удаленных NUMA-узлов, в L2/L3 кеш-памяти других ядер или процессора. В то же время иерархический алгоритм гарантирует локализацию взаимодействий через локальную (общую) для группы процессов L2/L3 кеш-память, контроллер памяти NUMA-узла. На рис. 6 показано время выполнения алгоритмов для трех схем назначения процессов на ядра. Иерархический алгоритм демонстрирует хорошую устойчивость к изменению способа распределения процессов – коэффициент вариации времени выполнения алгоритма не превышает 7 % для трех рассмотренных схем, при этом алгоритм обеспечивает минимум времени реализации барьерной синхронизации на данной системе.

Алгоритм MPI_Barrier	Времы выг распре	Коэффициент вариации времени выполнения (%)		
	map-by core			
shm-sense-reversing	50,35	45,55	52,8	7,4
shm-sense-reversing-counter	4,41	5,1	4,64	7,4
shm-sense-reversing-gr	6,80	7,18	6,98	2,7
shm-combining-tree	3,09	3,92	3,63	11,9
shm-mcs	3,31	5,47	4,9	24,5
shm-tournament	3,83	3,59	7,38	43,0
shm-dissemination	4,37	3,45	3,85	11,9
shm-topo-NUMA-SOCKET	3,25	2,86	3,16	6,6
shm-topo-NUMA	2,98	2,68	3,03	6,5
shm-topo-SOCKET	3,69	3,34	3,16	7,9

Рис. 6. Время выполнения алгоритмов операции MPI_Barrier для трех схем распределения процессов по 128 ядрам двух процессоров (4 NUMA-узла по 32 ядра)

4. Заключение

Разработанный иерархический алгоритм барьерной синхронизации обеспечивает многоуровневую синхронизацию между процессами, объединёнными в группы в соответствии с разделяемыми ими уровнями иерархии памяти (кеш-память L2/L3, NUMA-узел, процессор). Алгоритм реализован на базе библиотеки Open MPI. Эксперименты на сервере с двумя процессорами Huawei Kunpeng (128 ядер, 4 NUMA-узла) показали, что разработанный алгоритм с группировкой процессов по NUMA-узлам обеспечивает минимальное время выполнения по сравнению с реализованными в библиотеке Open MPI и устойчив к изменению способа распределения процессов по процессорным ядрам.

Литература

- Jain S., Kaleem R., Balmana M., Langer A., Durnov D., Sannikov A. and Garzaran M. Framework for Scalable Intra-Node Collective Operations using Shared Memory // Proc. of the International Conference for High Performance Computing, Networking, Storage, and Analysis (SC-2018), 2018. P. 374–385.
- 2. *Курносов М. Г., Токмашева Е. И.* Оптимизация барьерной синхронизации на асимметричных NUMA-подсистемах процессорных ядер // Вестник СибГУТИ. 2021. № 1. С. 36–49.
- 3. Yew P. C., Tzeng N. F., Lawrie D. H. Distributing Hot Spot Addressing in Large Scale Multiprocessors // IEEE Transactions on Computers. 1987. V. C-36, Is. 4. P. 388–395.
- 4. *Mellor-Crummey J. M., Scott M. L.* Algorithms for Scalable Synchronization on Shared-memory Multiprocessors // ACM Transactions on Computer Systems. 1991. № 9 (1). P. 21–65.
- 5. *Hengsen D., Finkel R., Manber U.* Two Algorithms for Barrier Synchronization // Int. Journal of Parallel Programming. 1988. V. 17, Is. 1. P. 1–17.
- 6. Brooks E. The butterfly barrier // Journal of Parallel Programming. 1986. V. 15, Is. 4. P. 295–307.

Статья поступила в редакцию 21.03.2022.

Курносов Михаил Георгиевич

д.т.н., профессор, профессор кафедры вычислительных систем Сибирского государственного университета телекоммуникаций и информатики (630102, Новосибирск, ул. Кирова, 86), тел. (383) 269-83-82, e-mail: mkurnosov@sibguti.ru; старший научный сотрудник Института физики полупроводников им. А. В. Ржанова СО РАН (630090, Новосибирск, просп. Академика Лаврентьева, 13), тел. (383) 330-56-26, e-mail: mkurnosov@isp.nsc.ru.

Barrier synchronization hierarchical algorithm for multicore shared-memory systems

Mikhail G. Kurnosov

Doctor of technical sciences, Professor, Siberian State University of Telecommunications and Information Science (SibSUTIS, Novosibirsk, Russia), mkurnosov@sibguti.ru.

A hierarchical MPI barrier synchronization algorithm creating groups of processes that share common resources at the memory hierarchy levels (L2/L3 caches, NUMA node, socket) is proposed. Synchronization is performed in groups at each level of the hierarchy. Experiments on a dual-socket server with two Huawei Kunpeng processors (128 cores, 4 NUMA nodes) showed that the proposed algorithm with NUMA nodes process grouping provides the minimum execution time compared to known methods and is resistant to different schemes of process placement.

Keywords: barrier, shared memory, MPI, NUMA.

References

- 1. Jain S., Kaleem R., Balmana M., Langer A., Durnov D., Sannikov A. and Garzaran M. Framework for Scalable Intra-Node Collective Operations using Shared Memory. *Proc. of the International Conference for High Performance Computing, Networking, Storage, and Analysis (SC-2018)*, 2018, pp. 374-385.
- Kurnosov M.G., Tokmasheva E.I. Optimizacija bar'ernoj sinhronizacii na asimmetrichnyh NUMApodsistemah processornyh jader [Barrier Optimization on Asymmetrical NUMA Subsystems]. *Vestnik SibGUTI*, 2021, no. 1, pp. 36-49.
- 3. Yew P.C., Tzeng N.F., Lawrie D.H. Distributing Hot Spot Addressing in Large Scale Multiprocessors. *IEEE Transactions on Computers*, 1987, vol. C-36, iss. 4, pp. 388-395.
- 4. Mellor-Crummey J.M., Scott M.L. Algorithms for Scalable Synchronization on Shared-memory Multiprocessors. *ACM Transactions on Computer Systems*, 1991, vol. 9(1), pp. 21-65.
- 5. Hengsen D., Finkel R., Manber U. Two Algorithms for Barrier Synchronization. *Int. Journal of Parallel Programming*, 1988, vol. 17, iss. 1, pp. 1-17.
- 6. Brooks E. The butterfly barrier. Journal of Parallel Programming, 1986, vol. 15, iss. 4, pp. 295-307.

Применение специализированных детерминированных конечных автоматов для прогнозирования временных рядов

У. В. Павлова, А. А. Ракитский

В данной работе рассматривается возможность использования детерминированных конечных автоматов для прогнозирования временных рядов в режиме реального времени. Для решения этой проблемы авторами предлагается модификация автомата с десятью считывающими головками, предназначенного для распознавания мультилинейных последовательностей. В статье представлены модификации этого автомата, алгоритмы для их реализаций и приведены результаты, полученные с их помощью. Кроме того, в работе представлен метод изменения алгоритма работы автомата, обеспечивающий его пошаговое выполнение.

Ключевые слова: конечный автомат, прогнозирование, анализ временных рядов, теория информации, машинное обучение, сжатие данных.

1. Введение

В настоящее время исследования в области анализа и прогнозирования временных рядов охватывают практически все сферы деятельности человека: прогнозирование сбоев в работе различного оборудования; анализ экономических показателей; исследования различных данных в области медицины, таких как электрокардиограммы, электроэнцефалограммы и другие; природные явления, среди которых прогнозирование появления солнечных пятен, сейсмической активности, погодных явлений и многое другое. Все описанные выше примеры представляют собой временные ряды – последовательности данных, распределённых во времени (как правило, равномерно). В настоящее время большая часть данных записывается в режиме реального времени и имеет огромные объемы, в связи с чем необходимы надежные и эффективные средства для их обработки и анализа. К сожалению, большинство современных достаточно точных методов требуют огромных вычислительных ресурсов, поэтому разработка новых, высокопроизводительных, методов, позволяющих осуществлять анализ временных рядов в режиме онлайн, становится крайне актуальной.

В рамках представленной работы предлагается именно такой высокопроизводительный метод, позволяющий выявлять сложные паттерны во временных рядах в режиме реального времени. Предлагаемый метод базируется на конечном детерминированном автомате с десятью считывающими головками, который предназначен для распознавания мультилинейных последовательностей [11]. В работе исследуется возможность формирования таких конечных автоматов, которые смогли бы определять последовательность, выдаваемую неизвестным источником. Авторами рассмотрены несколько модификаций описанного выше автомата, позволяющих, во-первых, обрабатывать последовательности, отличающиеся от мультилинейных, а во-вторых, работать в режиме реального времени. Кроме того, рассмотрена возможность работы в сочетании с другими, уже известными, методами, для уточнения прогнозируемого значения. В последующих главах представлен обзор существующих методов прогнозирования,

описана модель автомата, представлены алгоритмы методов, предложенных авторами, а также результаты применения разработанного метода для некоторых реальных временных рядов.

2. Анализ предметной области

В работе [1] представлена модель, включающая в себя основные методы, используемые в машинном обучении, такие как искусственная нейронная сеть, многомерные адаптивные регрессионные сплайны, метод k-ближайших соседей и векторная регрессия опорных векторов. Авторы описывают применение разработанной модели к прогнозированию количества осадков в городах, а также приводят анализ эффективности своего алгоритма с входящими в его состав методами. В работе [2] описан метод пространственного прогнозирования электрической нагрузки с использованием модели клеточного автомата для пространственно-временного распределения новых нагрузок в определенной зоне обслуживания. Плотность электрической нагрузки для каждого из основных классов потребителей в каждой ячейке используется в качестве текущего состояния. Дополнительно устанавливается ряд правил обновления для имитации поведения роста и взаимодополняемости между классами. Положительные черты, выделяемые у этого метода, – хорошая производительность, небольшой объем необходимых данных и простота самого алгоритма, что позволяет масштабировать его без существенного увеличения сложности.

В [3] авторы описывают новый конечный обучающий автомат, который был разработан для снижения энергозатрат при преобразовании автоматов в детерминированные. Новый алгоритм содержит особый параметр для нахождения компромисса между точностью и потреблением энергии, что позволяет упростить машинное обучение. Благодаря своей детерминированности такой автомат может ускорить рабочий процесс и точно найти следующий символ в последовательности, используя небольшое количество операций, обеспечивая тем самым максимально быстрое прогнозирование.

Авторы статьи [4] разработали алгоритм Wayeb – инструмент, который пытается решить проблему прогнозирования сложных событий. Он использует автомат в вычислительной модели для обнаружения шаблонов и цепей Маркова, чтобы обеспечить вероятностное описание автомата. В статье [5] описаны регрессионные автоматы, которые предоставляют новый тип синтаксической модели для прогнозирования временных рядов. Построенный на основе обычных алгоритмов слияния состояний для идентификации автоматов, этот метод использует числовые данные в дополнение к символьным значениям и делает прогнозы на основе этих данных в режиме регрессии. Эти модели применяются для почасового прогнозирования скорости и энергии ветра. Также необходимо упомянуть возможность использования цепей Маркова [4, 6] и широко распространенного алгоритма ARIMA [7] для задач прогнозирования временных рядов.

В работах [8, 9] показано, что для прогнозирования временных рядов также могут быть использованы модели, основанные на использовании универсальных кодов (или методов сжатия данных). Так как архиваторы по своей структуре близки к автоматам и включают в себя алгоритмы сжатия данных, то их использование в задачах прогнозирования возможно как в качестве самостоятельного элемента, так и в комплексе с другими методами.

Конечные автоматы в анализе временных рядов – малоизученное направление. Возможность их применения только исследуется учеными. Конечный автомат способен не только распознавать слова, состоящие из символов теоретически бесконечного алфавита, но и при правильной реализации его можно обучить таким образом, чтобы он распознавал, к какому конкретно языку относится то или иное слово. В данной работе предлагается применение конечных автоматов для анализа и прогнозирования временных рядов, в которых могут присутствовать так называемые выбросы, т.е. отклонения от некоторого паттерна. К таким временным рядам, как правило, относятся реальные последовательности с неизвестными свойствами.

3. Теоретический базис

3.1. Задача прогнозирования

Прогнозирование является распространённой и востребованной задачей во многих областях человеческой деятельности. В результате использования методов прогнозирования уменьшается риск принятия неверных, необоснованных или субъективных решений. В целом задачу прогнозирования можно считать одной из самых сложных среди существующих задач. Она включает в себя множество неупорядоченных наборов данных, которые описывают поведение того или иного ряда. Для того чтобы составить прогноз, необходимо установить зависимость между всеми наборами, участвующими в прогнозе, а также проанализировать состояние каждого из них в прошлом и настоящем. Таким образом, можно сказать, что задача прогнозирования требует наличия некоторой выборки для обучения.

Очевидно, что данные, поступающие для обработки и дальнейшего прогнозирования, могут иметь сложную структуру, а их источник может быть недетерминированным или вероятностным и, кроме того, может быть чувствительным к внешним воздействиям. Если не накладывать на задачу какие-либо ограничения, упрощающие работу, то её можно отнести к разряду практически нерешаемых. В рамках работы будем считать, что источник данных полностью детерминирован, имеет конечный алфавит и генерирует не более одного символа в каждый момент времени. При таких упрощениях задача сводится к прогнозированию последовательности бесконечной длины.

3.2. Автоматы

Согласно определению, предложенному в [10], конечный автомат – абстрактная машина, число возможных внутренних состояний которой конечно.

Определим некоторый автомат М:

$$M = (V, Q, q_0, F, \delta),$$

где *V* – входной алфавит (конечное множество входных символов), из которого формируются входные слова, воспринимаемые конечным автоматом;

Q – множество внутренних состояний;

 q_0 – начальное состояние ($q_0 \in Q$);

F – множество заключительных, или конечных, состояний;

 δ – функция переходов, значение которой есть множество всех состояний, в которые из данного состояния возможен переход по конкретному входному символу или пустой цепочке (ϵ).

Конечный автомат начинает работу в состоянии q_0 , последовательно считывая по одному символу входного слова либо подцепочки символов этого слова (которые также можно рассмотреть как один символ, если провести преобразование алфавита). После прочтения каждого символа автомат переходит в новое состояние в соответствии с функцией переходов.

В рассматриваемой модели каждый автомат M принимает на вход бесконечное слово α и выдаёт на выходе бесконечное слово $M(\alpha)$ с таким ограничением, что, прежде чем считать очередной *i*-й символ, автомат должен вывести M(i), то есть спрогнозировать следующий символ в слове. Если каждый *i*-й символ $M(\alpha)$ равен *i*-му символу α , то можно сказать, что автомат M полностью освоил цепочку α . В случае, когда некоторый автомат M сможет полностью распознать последовательность $\alpha(A)$ – слово над A, где A – конечный алфавит, то можно утверждать, что автомат полностью освоил эту последовательность.

В дальнейшем автомату *М* подается на вход бесконечное слово *а*. На основе первого считанного символа и в соответствии со своим текущим состоянием автомат начинает делать переходы. При каждом движении он перемещает считывающую головку вправо и делает предположение о том, каким будет следующий символ во входящей последовательности. Общая схема работы детерминированного конечного автомата представлена на рис. 1.



Входная лента: входные символы а $_{i}\in\Sigma$

Рис. 1. Детерминированный конечный автомат

3.3. Автомат для распознавания мультилинейных последовательностей

Оригинальный автомат был представлен автором в виде псевдокода в работе [11], что оставляет пространство для различных его реализаций. Кратко опишем основные принципы работы алгоритма.

Любую бесконечную мультилинейную последовательность можно представить в виде:

$$q \prod_{n \ge 1} \prod_{i \ge 1}^{m} p_i s_i^n \tag{1}$$

для некоторых $m \ge 1$ и строк q, p_i , s_i , разделенных на блоки типа $p_i s_i^n$. То есть некоторая последовательность α может быть разбита на блоки, которые состоят из сегментов типа $p_i s_i^n$ и начального блока q. Для того чтобы полностью освоить a, автомат M будет чередовать в своей работе две процедуры – Коррекция и Сопоставление. Коррекция пытается расставить головки h_1 , h_2 , h_3 , h_4 в начало каждого сегмента. Более того, h_2 должна находиться после h_1 на заданном количестве сегментов (длина сегментов может отличаться), h_3 должна находиться после h_2 на таком же количестве сегментов и h_4 должна находиться после h_3 на таком же количестве сегментов. Каждый раз, когда запускается Коррекция, количество символов между головками увеличивается на 1. Сопоставление пытается освоить а, предполагая, что Коррекция была успешной и количество сегментов, разделяющих головки, одинаково для каждой из них. Сопоставление работает, используя h_1, h_2 , и h_3 для прогнозирования h_4 . Более «поздние» головки обозревают большее количество копий сегментов $p_i s_i^n$, то есть h_4 обозревает больше копий, чем h_3 , которая обозревает больше копий, чем h_2 , которая обозревает больше копий, чем h_1 . Сопоставление сдвигает все головки одновременно, сравнивая их между собой и определяя, в какой момент каждая из головок достигла конца своего сегмента. Как только все головки достигнут конца своих сегментов, Сопоставление начинается заново. Если обнаружена какая-либо проблема, Сопоставление завершается, и Коррекция запускается снова.

На рис. 2 показаны первые три шага процедуры Коррекция. Здесь видно, как автомат пытается расположить все головки таким образом, чтобы существующая последовательность разбивалась на блоки.

h2,	Inner,	1	h1										
h3,	outer,												
h3a	r, t,												
	h4												
¥	•	+	+										
start	1	2	3	4	5	6	7	8	9	10	11	12	13
	а	b	а	а	b	а	а	а	b	а	а	а	а
													Step 1
h3,	h4	Inner,	Outer,	t			h2		1			1	
h3a		1	r, h1										
+	♦	¥	♦	+			+						
start	1	2	3	4	5	6	7	8	9	10	11	12	13
	а	b	а	a	b	а	a	а	b	а	а	a	a
													Step 2
h3a	h4		h1	Inner,	Outer,		h2	t				h3	
				1	r								
+	+		+	+	+		+	+				+	
start	1	2	3	4	5	6	7	8	9	10	11	12	13
	а	b	а	а	b	а	a	а	b	a	а	a	а
	· · · · ·				•	•		•		•	•	•	Sten 3

Рис. 2. Схема работы процедуры Коррекция

На рис. З показана работа процедуры Сопоставления после завершения Коррекции. Все считывающие головки расположены в начале блоков и готовы к работе. Первый шаг показывает начальное положение головок, а второй шаг показывает положение после перехода к следующему символу. Как только первая головка (h_1) достигает нового символа «b», Сопоставление завершается.



Рис. 3. Схема работы процедуры Сопоставление

4. Алгоритм пошагового исполнения

В процессе исследования была разработана модель [12], максимально похожая на описанную ранее. К ней было применено несколько модификаций, подробно описанных в работах [13, 14], а также интегрирована работа с архиваторами [15]. Результаты тестирования, приведенные в работах, коренным образом не меняют результаты, полученные при первичном тестировании исходной модели автомата. Таким образом, принято решение об изменении конструкции и принципа работы автомата, но при этом сохранении логики его работы.

В ходе предыдущей работы над моделями автомата возникала проблема невозможности остановки алгоритма в конкретный момент времени и внесения изменений в его работу. Новый алгоритм реализован с возможностью модели останавливаться и запоминать свое положение каждый раз после того, как крайняя правая головка была перемещена и был сделан прогноз, а также начинать работу с места последнего «сохранения», а не с начала последовательности, как это предполагается в исходной версии автомата.

Поскольку принципы машинного обучения должны быть соблюдены, необходимо разделить входную последовательность на обучающую и тестовую выборки. Из этого следует, что и работа автомата должна быть разграничена на зоны ответственности по обработке последовательности. В представленном алгоритме было введено два режима – обучения и генерации. Ниже каждый из них кратко описан с теоретической точки зрения.

В режиме обучения из последовательности, поступившей автомату в качестве входных данных, выделяется и сохраняется паттерн, исходя из которого будет проводиться дальнейшая работа алгоритма. Стоит отметить, что размер обучающей выборки ограничен определенным количеством символов и не может превышать заданного фиксированного значения. После успешного завершения режима обучения запускается режим генерации. В этой ветке работы автомат, исходя из выбранного ранее шаблона, определяет следующий символ в последовательности. Далее подробно рассмотрим автомат и его работу на примере псевдокода.

```
Input: Sequence, data_size, learn_size
Output: Automaton result
   create multiHeadAutomaton A (sequence, data_size)
   If (A->learn(learn_size)) success:
        A->generateNext (generate_size)
```

Алгоритм 1. Маіп-функция

В main-функции программы (см. Алгоритм 1) на вход поступает последовательность, с которой будет происходить дальнейшая работа, её размер и длина обучающей выборки. В процессе создания экземпляра класса конструктор (Алгоритм 2) принимает непосредственно строку с данными и её длину. После создания по отношению к созданному экземпляру запускается процесс обучения, а в случае его успешного завершения – процесс генерации.

```
class multiHeadAutomaton:
    data_type data
    vector <data_type> predict_arr
    data_size, match_state, h1, h2, h3, h3a, h4, t, l, r, inner, outer, corr
    pattern_check = 0
    multiHeadAutomaton(data_type *d, int dsize):data(d),data_size(dsize)
```

Алгоритм 2. Поля и конструктор класса

Перейдем к более детальному описанию класса автомата. Полями являются считывающие (h_1, h_2, h_3, h_4) и вспомогательные (остальные) головки автомата, с помощью которых происходит основная работа, а также дополнительные переменные, необходимые в процессе выполнения алгоритма. Кроме того, как показано в Алгоритме 1, класс содержит конструктор и несколько методов, которые необходимы для его работы. В рамках данной работы подробно будут рассмотрены основные из них. Подробное описание остальных методов можно найти в работах [11, 12].

Прежде чем запустить один из режимов, автомат считывает входную последовательность и размер её обучающей выборки. Далее запускается непосредственно режим обучения. Он начинается с вызова метода *learn()* (см. Алгоритм 3), на вход которому поступает размер обучающей выборки. Внутри себя функция вызывает два ключевых метода – *Коррекцию* и *Conocmaвление*.

Метод Коррекция сохранил свой первоначальный вид и не претерпел изменений в структуре. Внутри него автомат расставляет считывающие головки по последовательности и с помощью них, а также вспомогательных выделяет паттерн. Если паттерн выделился успешно, то он проверяется в процедуре *Сопоставления*. В том случае, если вернулся 0, паттерн считается ошибочным, и Коррекция запускается снова. Если шаблон не был выделен, то размер обучающей выборки был слишком мал либо последовательность содержит много стохастических выбросов. В таком случае за паттерн принимается последняя найденная подстрока.

```
Input: lsize
Output: Learn result
while h4 < lsize:
   Correction()
   while Matching(lsize)==0</pre>
```

Алгоритм 3. Функция learn()

После завершения функции *Коррекция* вызывается функция *Сопоставление* (Алгоритм 4), в которой происходит основная работа и проверяется правильность определенного паттерна. Автомат сравнивает каждый символ выбранного шаблона на равенство соответствующему символу в обрабатываемой последовательности.

```
Input: Automaton state (h1, h2, h3, h3a, h4, r, l, t, inner, outer), size
Output: Matching result
Loop:
If match_state <= 0:</pre>
   h3a := h3
   match_state := 0
 If match state <= 1:</pre>
   If h4 > size:
return 1
   match_state := 1
   If step_one():
     return 0
   If data[h2] <> data[h4]:
     return 1
 If match_state <= 4:</pre>
   If h4 > size:
     return 1
   match state := 4
   If step_four():
     return 0
   match state := 0
   If data[h3a] <> data[h4]:
     data[h4] := data[h3a]
     h4 + 1
     h3a + 1
End loop
```

Алгоритм 4. Функция Сопоставление

Сначала проверяется текущее состояние автомата. В момент, когда автомат определил свое фактическое состояние, в зависимости от его значения он запускает одну из внутренних функций *step()* (Алгоритм 5). Эти функции перемещают считывающие головки и вызывают функцию прогнозирования. В ней автомат сравнивает предсказанный символ с символом, содержащимся в исходной последовательности. Если символы равны, это означает, что нужно сохранить символ и рассматривать его как часть паттерна. В противном случае найденный паттерн неверен, поэтому автомат прекращает *Conocmaвление* и снова запускает *Коррекцию*.

h3a + 1			
h3 + 1			
h4 + 1			
Guess(h2, h	ו4)		

Алгоритм 5. Функция step()

Если правильный паттерн был успешно выделен, автомат должен прекратить вызывать функцию Коррекции. Когда это событие произошло, мы можем завершить обучение и запустить режим генерации, в котором обязательно необходимо указать, сколько символов последовательности нужно сгенерировать. В противном случае автомат сгенерирует один символ и прекратит работу. В результате внесенных изменений, представленных выше, мы получили автомат, полностью соответствующий исходному по логике выполнения.

5. Результаты тестирования

Для того чтобы проанализировать работу автомата, был проведен ряд тестов с использованием различных последовательностей. В данной работе рассматривается временной ряд с алфавитом {0, 1}, поэтому для оценки точности прогноза не принимается во внимание математическое ожидание и дисперсия, но ведется подсчет ошибок, совершенных автоматом во время работы. Под верным прогнозом будем понимать значение, которое спрогнозировал автомат, совпадающее со значением в исходной последовательности.

Подробно о предыдущих результатах говорилось в работах [12, 15]. Для дополнительного исследования возможностей автомата было проведено тестирование на псевдомультилинейной последовательности [16]. В исходном виде данная строка не имеет паттерна, поэтому автомат, как и ожидалось, не смог обработать строку и совершить какой-либо прогноз. Очевидно, что при данном результате последовательность должна подвергнуться предварительной обработке. В её начало вручную был добавлен паттерн. График обработанной последовательности отображен на рис. 4. На графике можно увидеть, каким образом изменяется последовательность в зависимости от количества считанных данных.



Рис. 4. Псевдомультилинейная последовательность

В ходе тестирования длина паттерна несколько раз была изменена. Это дало возможность оценить работу автомата при избыточной и недостаточной длине паттерна. Результаты привелены в табл. 1.

Длина последовательности	Размер обучающей выборки	Процент верных прогнозов
1000000	100000	86 %
1000000	10000	32 %
1000000	300000	65 %

Таблица 1. Результаты тестирования

Длина последовательности оставалась неизменной на протяжении всего тестирования. Первоначальный размер паттерна равнялся 10 % длины строки. В режиме обучения автомат выделил закономерность и перешел в режим генерации. Сравнение сгенерированной и первоначальной строки показало, что точность сделанных прогнозов достаточно высока для последовательности, не являющейся мультилинейной.

Далее размер обучающей выборки был сильно сокращен, что повлекло за собой ухудшение качества прогноза. Это объясняется особенностями работы автомата – в режиме обучения он успевает полностью выделить паттерн, что вызывает дальнейшие несовпадения при сравнении сгенерированной последовательности с оригинальной.

При увеличении длины паттерна процент верно сделанных предположений снижается, хоть и некритично. Автомат полностью выделяет не только сам паттерн, но и часть последовательности, что и вносит погрешности в дальнейшую работу.

6. Заключение

В статье рассматривается возможность применения детерминированных конечных автоматов в задаче прогнозирования временных рядов. Был рассмотрен десятиголовочный автомат, настроенный на работу с мультилинейными последовательностями, а также описаны его недостатки при работе с временными рядами, отклоняющимися от паттерна. Основное внимание уделено совершенно новому алгоритму, способному работать в пошаговом режиме. Алгоритм включает несколько режимов работы – обучение и прогнозирование. Каждый из этих режимов выполняет определённую функцию: обучение настраивает автомат и выделяет паттерн, а прогнозирование генерирует последовательность и сравнивает ее с исходной. Приведены результаты тестирования разработанного метода на последовательностях, близких к мультилинейным.

Литература

- 1. Sumi S. M., Zaman M. F., Hirose H. A rainfall forecasting method using machine learning models and its application to the Fukuoka city case // International Journal of Applied Mathematics and Computer Science. 2012. V. 22, № 4. P. 841–854.
- 2. Carreno E. M., Rocha R. M., Padilha-Feltrin A. A cellular automaton approach to spatial electric load forecasting // IEEE Transactions on Power Systems. 2010. V. 26, №. 2. P. 532–540.
- 3. *Abeyrathna K. D.* A Novel Multi-step Finite-State Automaton for Arbitrarily Deterministic Tsetlin Machine Learning // International Conference on Innovative Techniques and Applications of Artificial Intelligence. Cambridge, UK, December 15–17, 2020. V. 12498. P. 108.
- 4. *Alevizos E., Artikis A., Paliouras G.* Wayeb: a tool for complex event forecasting // 22nd International Conference on Logic for Programming, Artificial Intelligence and Reasoning. Awassa, Ethiopia, November 16–21, 2018.
- Qin L., Hammerschmidt C., Pellegrino G., Verwer S. Short-term Time Series Forecasting with Regression Automata // ACM SIGKDD 2016 Workshop on Mining and Learning from Time Series (MiLeTS). San Francisco, California, USA, August 13–17, 2016. DOI: 10.1145/1235.

- 6. *Alevizos E., Artikis A., Paliouras G.* Event Forecasting with Pattern Markov Chains // 11th ACM Int. Conf. on Distributed and Event-based Systems, 2017. P. 146–157.
- 7. Jenkins G. M., Box G. Time Series Analyses. Forecasting and control. Holden-Day, 1976. 589 p.
- 8. *Lysyak A. S., Ryabko B. Y.* Time series prediction based on data compression methods // Problems of Information Transmission. 2016. V. 52, № 1. P. 92–99.
- 9. *Чирихин К. С., Рябко Б. Я.* Экспериментальное исследование точности методов прогноза, базирующихся на архиваторах // Вестник Новосибирского государственного университета. Серия: Информационные технологии. 2018. Т. 16, №. 3. С. 145–168.
- 10. *Hopcroft J. E., Ullman J. D.* Introduction to automata theory, languages, computation. Addison-Wesley publishing company, 1979. 427 p.
- 11. *Smith T.* Prediction of infinite words with automata // Theory of Computing Systems. 2018. V. 62, № 3. P. 653–681.
- 12. Павлова У. В., Ракитский А. А. Исследование возможности применения автоматов для прогнозирования временных рядов // РНТК «Обработка информации и математическое моделирование», СибГУТИ, 25–26 апреля 2019. С. 168–172.
- Pavlova U., Rakitskiy A. Time Series Forecasting Method Based on Finite State Machine // International Conference of Young Specialists on Micro/Nanotechnologies and Electron Devices (EDM). Aya, Altai Region, Russia, June 30 July 3, 2021. P. 533–536.
- Pavlova U., Rakitskiy A. Development and Research of the Time Series Prediction Method Based on Finite State Automaton // Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT), Yekaterinburg, Russia, 13–14 May, 2021. P. 305–307.
- 15. Павлова У. В., Ракитский А. А. Разработка и исследование метода прогнозирования временных рядов на основе конечных автоматов // РНТК «Обработка информации и математическое моделирование», СибГУТИ, 23–24 апреля 2020. С. 142–145.
- 16. UCI Machine Learning Repository [Электронный pecypc]. URL: https://archive.ics.uci.edu/ml/index.php (дата обращения: 28.01.2022).

Статья поступила в редакцию 18.03.2022; переработанный вариант – 05.04.2022.

Павлова Ульяна Владимировна

аспирант 1-го курса института ИВТ СибГУТИ, ассистент кафедры ПМиК СибГУТИ (630102, Новосибирск, ул. Кирова, 86), e-mail: uljana.pavlova2012@yandex.ru.

Ракитский Антон Андреевич

к.т.н, доцент кафедры ПМиК (630102, Новосибирск, ул. Кирова, 86); научный сотрудник НГУ; старший научный сотрудник ИВТ СО РАН, e-mail: rakitsky.anton@gmail.com.

Application of specialized DFA for time series forecasting

Ulyana V. Pavlova

Postgraduate student, Department of Appl. Math. and Cybernetics, Siberian State University of Telecommunications and Information Science (SibSUTIS, Novosibirsk, Russia), uljana.pavlova2012@yandex.ru.

Anton A. Rakitskiy

Candidate of technical sciences, Docent, Siberian State University of Telecommunications and Information Science (SibSUTIS, Novosibirsk, Russia), rakitsky.anton@gmail.com.

In this paper the possibility of using deterministic finite automaton to forecast time series in real time is considered. To solve this problem, a modification of the automaton with 10 reading heads designed to recognize multilinear sequences is proposed. The article presents modifications of this automaton, algorithms for their implementation, and demonstrates the results for various time series. In addition, the paper presents a method for changing the algorithm of the automaton ensuring its step-by-step execution.

Keywords: finite state automaton, forecasting, time series analysis, information theory, machine learning, data compression.

References

- 1. Sumi S. M., Zaman M. F., Hirose H. A rainfall forecasting method using machine learning models and its application to the Fukuoka city case. *International Journal of Applied Mathematics and Computer Science*, 2012, vol. 22, no. 4, pp. 841-854. DOI: 10.2478/v10006-012-0062-1.
- Carreno E. M., Rocha R. M., Padilha-Feltrin A. A cellular automaton approach to spatial electric load forecasting. *IEEE Transactions on Power Systems*, 2010, vol. 26, no. 2, pp. 532-540. DOI: 10.1109/tpwrs.2010.2061877.
- 3. Abeyrathna K. D. A Novel Multi-step Finite-State Automaton for Arbitrarily Deterministic Tsetlin Machine Learning. *International Conference on Innovative Techniques and Applications of Artificial Intelligence*, Cambridge, UK, 15-17 December, 2020, vol. 12498, pp. 108. DOI: 10.1007/978-3-030-63799-6 8.
- 4. Alevizos E., Artikis A., Paliouras G. Wayeb: a tool for complex event forecasting. 22nd International Conference on Logic for Programming, Artificial Intelligence and Reasoning, Awassa, Ethiopia, 16-21 November, 2018.
- 5. Qin L., Hammerschmidt C., Pellegrino G., Verwer S. Short-term Time Series Forecasting with Regression Automata. *ACM SIGKDD 2016 Workshop on Mining and Learning from Time Series (MiLeTS)*, San Francisco, California, USA, 13-17 August, 2016. DOI: 10.1145/1235.
- 6. Alevizos E., Artikis A., Paliouras G. Event Forecasting with Pattern Markov Chains. *11th ACM Int'l Conf. on Distributed and Event-based Systems*, 2017, pp. 146-157. DOI: 10.1145/3093742.3093920.
- 7. Jenkins G. M., Box G. Time Series Analyses. Forecasting and control. Holden-Day, 1976, 589 p.
- 8. Lysyak A. S., Ryabko B. Y. Time series prediction based on data compression methods. *Problems of Information Transmission*, 2016, vol. 52, no. 1, pp. 92-99. DOI: 10.1134/s0032946016010075.
- Chirikhin, K. S., and B. Ya. Ryabko. Jeksperimental'noe issledovanie tochnosti metodov prognoza, bazirujushhihsja na arhiva-torah [Experimental Study of the Accuracy of Compression-Based Forecasting Methods]. *Vestnik NSU, Series: Information Technologies*, 2018, vol. 16, no. 3, pp. 145-158. DOI: 10.25205/1818-7900-2018-16-3-145-158.
- 10. Hopcroft J.E., Ullman J.D. *Introduction to automata theory, languages, computation*. Addison-Wesley publishing company, 1979, 427 p.
- 11. Smith T. Prediction of infinite words with automata. *Theory of Computing Systems*, 2018, vol. 62, no. 3, pp. 653 681. DOI: 10.1007/s00224-016-9739-4.
- 12. Pavlova U. V., Rakitsky A. A. Issledovanie vozmozhnosti primenenija avtomatov dlja prognozirovanija vremennyh rjadov [Investigation of the possibility of using automata to predict time series]. *Rossijskaja nauchno-tehnicheskaja konferencija*, Novosibirsk, 25-26 April, 2019, pp. 168-172.
- Pavlova, U., Rakitskiy, A. Time Series Forecasting Method Based on Finite State Machine. 2021 IEEE 22nd International Conference of Young Professionals in Electron Devices and Materials (EDM), Aya, Altai Region, Russian Federation, 30 June 3 July, 2021, pp. 533-536. DOI: 10.1109/edm52169.2021.9507729.
- Pavlova, U., Rakitskiy, A. Development and Research of the Time Series Prediction Method Based on Finite State Automaton. 2021 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT), Yekaterinburg, Russian Federation, 13-14 May, 2021, pp. 305-307. DOI: 10.1109/usbereit51232.2021.9455056.
- 15. Pavlova U. V., Rakitsky A. A. Razrabotka i issledovanie metoda prognozirovanija vremennyh rjadov na osnove konechnyh avtomatov [Development and research of a time series forecasting method based on finite automata]. *Rossijskaja nauchno-tehnicheskaja konferencija*, Novosibirsk, 23-24 April, 2020, pp. 142-145.
- 16. UCI Machine Learning Repository, available at: https://archive.ics.uci.edu/ml/index.php (accessed 28.01.2022).

УДК 004.414, 004.418. DOI: 10.55648/1998-6920-2022-16-2-23-32

Исследование актуальных требований для разработки систем мониторинга и управления телекоммуникационной сети

В. О. Доценко, И. Е. Шевнина

Современный рынок телекоммуникаций диктует необходимость модернизации существующих требований к системам мониторинга и управления сетей передачи данных. Инфраструктура действующих телекоммуникационных сетей непрерывно усложняется и расширяется, в связи с этим современные продукты устранения неисправностей и контроля производительности сети должны быть оснащены соответствующим функционалом. Проделанная работа определяет актуальные требования, выдвигаемые к современным системам, которые необходимо учитывать при разработке программных продуктов мониторинга и управления. В данной статье исследуются функциональные особенности программных комплексов, использующихся на сетях операторов связи, а именно: ТЕОСО Netrac 7, СОВА компании САТЕЛ, ЕМС Smarts, HP Network Node Manager, Zabbix, Cacti, NOC Project. Инструментарием, используемым в работе, выступают сравнение и анализ. Тема исследования удовлетворяет критериям направления деятельности Министерства цифрового развития, связи и массовых коммуникаций Российской Федерации (Федеральный закон №204-ФЗ от 23 июня 2016 года «О внесении изменений в статьи 2 и 24 Федерального закона «О связи»).

Ключевые слова: системы мониторинга и управления, телекоммуникационная сеть, устранение неисправностей.

1. Введение

IT-инфраструктуры сетей современных операторов связи включают в себя тысячи систем и сетевых устройств различных вендоров с уникальным программным обеспечением. С учётом того, что каждый производитель разрабатывает и обслуживает преимущественно свои стандарты, мониторинг и администрирование таких систем становится сложнейшей задачей.

Во время работы сети могут происходить сбои питания, ошибки программного обеспечения систем, ошибки эксплуатации оборудования, повреждение линейных участков, приводящие к сбоям, неисправностям и авариям. С целью обеспечения стабильного функционирования сети ответственному IT-персоналу, обслуживающему сеть передачи данных, необходимо иметь удобный функционал по отслеживанию состояния оборудования и утилизации каналов, контролировать качество предоставляемого обслуживания пользователям, прогнозировать потенциальные неисправности, чётко понимать схему организации связи и иметь возможность быстрого устранения сетевых проблем любого характера.

Для выполнения поставленных задач стандартно используются системы устранения неисправностей (Fault Management, FM) и контроля производительности сети (Perfomance Management, PM). Они представляют собой системы контроля и управления аварийными сигналами, предназначенные для фильтрации лишней информации и корреляции актуальных событий с целью выявления первопричины, породившей поток взаимосвязанных аварийных сообщений. Применение таких систем мониторинга и управления (далее – СМУ) позволяет оптимально использовать имеющиеся ресурсы компании за счёт упрощения анализа состояния сети и сокращения времени на поиск неисправностей.

Неудивительно, что провайдеры всё чаще рассматривают выбор СМУ как вопрос стратегической важности, так как на современном конкурентном рынке связи лидируют поставщики услуг, которые предоставляют экономически выгодные, высококачественные цифровые услуги. Точный сбор, обработка и анализ информации о трафике и ключевых показателей эффективности неизменно приводят к повышению качества обслуживания пользователей.

С учётом непрерывного развития современных телекоммуникационных сетей усложняются способы мониторинга неисправностей и контроля эффективности их работы. По данной причине с целью создания нового программного обеспечения для устранения неисправностей и контроля производительности современных телекоммуникационных сетей необходимо учитывать актуальные требования, выдвигаемые к разрабатываемым СМУ.

На данный момент к применяемым СМУ выдвигается перечень основных требований, который необходим для эффективной работоспособности сети, а именно:

1) расширяемость: автоматическое обнаружение устройств в сети, формирование списков оборудования, реакция на изменения параметров сети;

2) распределённость: запись и хранение основных настроек сети с распределённых составных частей системы, ведение отчетов в режиме реального времени для последующего анализа системных данных и сетевых проблем, а также планирования изменений;

3) масштабируемость: работа системы на кластерах с числом процессоров, изменяющимся в широком диапазоне;

 переносимость: комплекс должен работать на open-source-системе (например, Linux) для совместимости с другими продуктами и неподверженности изменениям от сопутствующего программного обеспечения;

5) экономичность: работа системы не должна оказывать существенного влияния на функционирование сети, клиентских услуг и приложений [1, 2];

6) рабочая система оповещения: контроль доступности оборудования и утилизации каналов для выявления проблем в кратчайшие сроки, поиска корневой причины инцидента и передачи информации ответственному персоналу;

7) прогнозирование: возможность устранения вероятных проблем в работе сетевого оборудования за счёт моделирования и использования продвинутых алгоритмов;

8) удобство использования: наличие интуитивно-понятного интерфейса для управления сетью [3, 4].

Вышеперечисленные требования являются основополагающими для существующих СМУ, однако к текущему моменту на сетях операторов связи используются системы с расширенным функционалом, который значительно повышает эффективность мониторинга и администрирования сети. К таким системам можно отнести: TEOCO Netrac 7, COBA компании CATEЛ, EMC Smarts, HP Network Node Manager, Zabbix, Cacti, NOC Project. На основе исследования функциональных особенностей перечисленных продуктов предложено расширить основные требования, выдвигаемые к разрабатываемым СМУ.

2. Краткое описание систем мониторинга и управления, использующихся на сетях операторов связи

2.1. TEOCO Netrac 7

Netrac 7 – это интегрированное, кросс-доменное решение по обеспечению и анализу услуг для операторов связи, включает в себя более 100 функциональных возможностей и модулей, которые повышают качество потребительских услуг. Платформа предназначена для мониторинга, конфигурирования, анализа загрузки и администрирования ресурсов в соответствии со спецификациями и стандартами, предложенными некоммерческим консорциумом TeleManagement Forum. Netrac 7 позволяет полностью управлять гетерогенными сетями, услугами и потребительским качеством обслуживания во всех технологических доменах [5, 6].

2.2. Система обнаружения и визуализации аварий (СОВА)

Система обнаружения и визуализации аварий – это промышленная платформа на открытом коде, состоящая из взаимосвязанных элементов. СОВА обладает высокой степенью интеграции и упрощает эксплуатацию, планирование и развитие телекоммуникационной или иной инфраструктуры. Данная система мониторинга от компании САТЕЛ предназначена для автоматизации процессов обнаружения неисправностей, отслеживания изменений конфигураций оборудования, контроля состояния услуг и производительности сети [7].

2.3. EMC Smarts

Продукт Smarts компании Dell EMC является продвинутой системой мониторинга и управления, в состав которой входят следующие компоненты: система сбора, хранения и анализа аварийных сообщений (FM), а также система мониторинга производительности и контроля сервисных соглашений (PM). EMC Smarts в достаточной степени отвечает таким требованиям, как непрерывность и надежность функционирования бизнес-процессов, быстродействие и удобство использования инструментов IT-инфраструктуры [8, 9].

2.4. HP Network Node Manager (NNM)

HP Network Node Manager – это платформа мониторинга и управления телекоммуникационной сети, которая является основой для построения центра управления сети операторов связи. На текущий момент NNM занимает одну из ведущих позиций в сетевом администрировании, обеспечивает мониторинг доступности устройств сети и является основным решением для сетевого управления [10].

2.5. Zabbix

Zabbix – это универсальное средство мониторинга, способное следить за динамикой работы серверов и сетевого оборудования, оперативно реагировать на внештатные ситуации и предупреждать о возможных проблемах с нагрузкой. Данная система мониторинга может записывать статистику из указанной рабочей среды и действовать в определенных случаях требуемым образом [11].

2.6. Cacti

Cacti – это надежная система управления производительностью и неисправностями сети. Данная система является open-source веб-приложением, которое обеспечивает сбор и хранение всей необходимой информации о сети в базе данных за счёт применения различных удалённых сборщиков. Cacti позволяет отобразить все накопленные статистические данные о сети в графическом виде [12].

2.7. NOC Project

Центр управления сетью (NOC) – специализированная программная среда для работы с сетевыми устройствами, решающая текущие задачи функционирования сети. Данный программный комплекс осуществляет круглосуточный мониторинг и управление сетями, позволяет снижать аварийность, обеспечивать высокую производительность инфраструктуры, повышая эффективность предоставления услуг при одновременном снижении рисков. Главная задача NOC – обеспечение бесперебойной работы сетей передачи данных, систем хранения, вычислительных ресурсов и сервисов клиентов [13, 14].

3. Сравнительный анализ функциональных возможностей обозреваемых систем

С целью формулирования актуальных требований к разрабатываемым СМУ были выделены наиболее востребованные функции в применяемых СМУ. В табл. 1 представлен результат сравнения особенностей обозреваемых программных продуктов из пункта 2.

N⁰	Название функции	Netrac 7	COBA	Smarts	NNM	Zabbix	Cacti	NOC	Встречае- мость, %
1	Применение графических средств	+	+	+	+	+	+	+	100
2	Построение топологии сети	+	+	+	+	+	+	+	100
3	Автоматическое устранение неисправностей	+	+	+	+	+	-	+	85.7
4	Отслеживание качества обслу- живания пользо- вателей	+	-	+	+	+	-	+	71.4
5	Резервирование данных	-	+	+	+	+	+	-	71.4
6	VLAN Management	-	+	+	+	-	-	+	57.1
7	IP Address Management	-	+	+	+	+	-	-	57.1
8	Привязка к гео- графическим данным	+	+	-	-	+	-	+	57.1
9	Configuration Management	+	+	-	+	-	-	+	57.1
10	Машинное обучение	+	-	-	+	+	-	-	42.9
11	DNS Provisioning	-	+	+	-	-	-	+	42.9
12	Peering Management	-	+	-	+	-	-	+	42.9
13	Способность к самомонито- рингу	+	-	-	+	+	-	-	42.9
14	Адаптация поро- говых значений	+	-	-	-	+	-	-	28.6
15	Защита системы	-	-	-	+	+	-	-	28.6
16	Балансировка нагрузки	-	-	-	-	-	+	-	14.3

Таблица 1. Сравнение функционала современных СМУ операторов связи

Из табл. 1 можно выделить частоту реализации функционала в различных СМУ, что определяет важность используемых функций для персонала, применяющего СМУ на сети, поэтому было принято решение разделить общий перечень функций на две части: базовый функционал современных СМУ и дополнительный функционал, который может быть полезен при разработке будущих программных продуктов.

С учётом частоты реализации базовых функций в программных комплексах была осуществлена приоритизация и дополнительное разделение на группы. Классификация функций по группам обусловлена требованием к увеличению приоритета при разработке программных продуктов. Таким образом, в первую очередь будут реализованы и оптимизированы наиболее важные для эффективного мониторинга и управления функции разрабатываемых СМУ.

Ниже представлены базовые функции современных СМУ, частота встречаемости которых более 50 %, а также их разделение с учётом приоритетности.

1. Группа высокого приоритета:

 применение графических средств: предоставление информации о сбоях в различных графических форматах, например, круговые диаграммы, карты, графики, таблицы и т.д.;

 построение топологии и карт: автоматическое построение, валидация и отображение карт топологии инфраструктуры, построение логических карт сервисов, взаимодействия подсистем с последующим мониторингом состояния обнаруженного оборудования [15, 16];

– привязка к географическим данным: представление данных мониторинга неисправностей на иерархической графической карте, а также представление её фрагментов, визуально отображающих информацию об оборудовании и событиях [7, 17].

2. Группа среднего приоритета:

 Сonfiguration Management: отслеживание изменений, обновлений данных конфигурации, управление конфигурацией оборудования, включая типовые настройки, соблюдение основных требований, оповещение о нарушении заданных правил, что помогает поднять уровень безопасности, доступности и операционной эффективности сети [15];

– автоматическое управление сетью: полностью автоматизированные алгоритмы, которые способны проверять конфигурацию оборудования на предмет соответствия установленным политикам и принимать корректирующие действия по устранению критических или подозрительных ситуаций, например, удалённо управлять оборудованием, автоматически перезапускать соответствующие процессы, управлять облачными ресурсами за счёт использования ранее подготовленных скриптов или сценариев и т.д.;

– обеспечение SLA (Service Level Agreement): построение и контроль модели сервисов на сети, проверка соответствия состояния предоставляемых сервисов, отслеживание соответствия качества обслуживания и проверка данных биллинга за счёт сбора информации о производительности элементов, вычисления характеристик производительности всей инфраструктуры или отдельных компонент/приложений и проверки их соответствия определенным значениям [16–19];

– резервирование: повышение отказоустойчивости системы за счёт использования серверного оборудования с дублированием основных аппаратных компонент и использования дублированных каналов СМУ, резервное копирование важной информации о сети, применение удаленных сборщиков данных – кэширование информации о событиях и состоянии сети во время простоя СМУ с последующей передачей обновлений в основную базу данных, как только она вернется в рабочий режим [12, 20].

3. Группа низкого приоритета:

– IP Address Management: мониторинг IP-адресов и адресного пространства, определение диапазонов IP-адресов для сканирования сети, поддержка независимых адресных пространств в отдельных VRF (Virtual Routing and Forwarding instance), иерархичность выделения блоков адресов, функционирование при десятках тысяч выделенных блоков и сотнях тысяч адресов IPv4 и IPv6;

– VLAN Management: возможность редактирования базы виртуальных локальных сетей (Virtual Local Area Network, VLAN), автоматическое обновление VLAN на нужном оборудовании вне зависимости от вендора, мониторинг информации о VLAN оборудования [18].

Функционал, который стоит принять в рассмотрение при разработке СМУ, однако не всегда он необходим для мониторинга и администрирования современных телекоммуникационных сетей – частота встречаемости менее 50 %:

 функционирование системы самомониторинга: отслеживание состояния компонентов системы, оповещение о сбоях и неисправностях функциональных блоков СМУ;

адаптация пороговых значений: применение набора алгоритмов с самообучением, автоматически подбирающего рекомендуемые пороги для наиболее эффективного мониторинга производительности сети и качества услуг, что существенно сокращает количество требуемых пороговых правил и время развертывания новых служб;

 применение алгоритмов машинного обучения: выявление аномалий на основе анализа исторических данных в режиме реального времени для прогнозирования возможных сетевых проблем [19];

 обеспечение защиты информации: использование методов шифрования служебных данных, использование гибких периодов хранения для различных наборов данных [12, 17];

– Peering Management: управление и документирование пиринговых сетей и пиринговых сеансов с помощью системы управления взаимодействием с внешними сетями IP, поддержка функций BGP (Border Gateway Protocol), что позволяет хранить базу пиров и генерировать фильтры для BGP;

– DNS Provisioning: применение интерфейса для управления зонами DNS (Domain Name System), использование системы именования объектов управления DNS с целью корректного разрешения IP-адресов, от чего зависит стабильная работа большинства компонентов системы управления [18];

 балансировка нагрузки: использование служб анализа утилизации каналов и балансировки нагрузки на агрегирующих интерфейсах с целью оптимизации использования сетевых ресурсов, сокращения времени обслуживания запросов, а также обеспечения отказоустойчивости [12].

Указанную категорию функций следует рассматривать в тех случаях, когда СМУ обеспечена необходимым функционалом, то есть выполнены все актуальные требования, предъявляемые к современным СМУ. Эффективность использования дополнительных функций последней описанной категории на данный момент явно не подтверждена, так что это направление можно определить как перспективное для дальнейших исследований.

В результате проведённого анализа были определены наиболее приоритетные функции, которые рекомендуется добавить в основной список требований к разрабатываемым программным продуктам мониторинга и управления сети. Ниже представлен перечень актуальных требований к СМУ с расставлением приоритетов для реализации:

1) применение графических средств: построение диаграмм, графиков и карт взаимодействия оборудования для мониторинга состояния обнаруженного оборудования;

 построение топологии инфраструктуры: автоматическое построение и отображение топологии сети с возможностью детализации её элементов, визуально отображающих информацию о событиях;

 привязка к географическим данным: представление данных мониторинга неисправностей на географической подложке и предоставление доступа к любому из объектов мониторинга с помощью геокарты;

4) функционал Configuration Management: отслеживание обновлений данных конфигураций оборудования, возможность системы проверять конфигурацию оборудования на предмет соответствия установленным политикам;

5) автоматическое устранение неисправностей: автоматическое управление конфигурацией, использование алгоритмов для проактивного выполнения прописанных сценариев и скриптов, перезапуска процессов;

6) отслеживание качества обслуживания пользователей: инструменты для проверки соответствия состояния предоставляемых сервисов установленным соглашениям;

7) резервирование системы: дублирование основных аппаратных компонентов, дублирование каналов СМУ, резервное копирование важной информации;

8) функционал IP Address Management: определение диапазонов IP-адресов для сканирования сети, мониторинг адресного пространства;

9) функционал VLAN Management: мониторинг информации о VLAN, автоматическое обновление базы VLAN и возможность её редактирования.

4. Заключение

По итогам исследования были обозначены расширенные требования CMУ, которые описывают рекомендуемый функционал для систем, разрабатываемых и применяемых для мониторинга и управления на современных телекоммуникационных сетях. Данные требования определяются базовой группой функций, полученных в ходе анализа вышепредставленных CMУ, так как их реализация обеспечивает удобство использования и отказоустойчивость системы, позволяет уменьшить время простоя услуг и контролировать заверенное качество обслуживания клиентов, таким образом повышая эффективность работы сети в целом.

Предлагаемые требования рекомендуется учитывать разработчикам ПО при создании будущих программных комплексов для мониторинга и управления сети. Разработка программных продуктов с учётом выведенных требований к системе позволит реализуемой СМУ конкурировать с текущими лидерами на рынке телекоммуникаций, а оптимизация и модернизация применяемого функционала повысит эффективность устранения неисправностей, контроля производительности и администрирования сетей.

Литература

- 1. Фокин В. Г. Управление телекоммуникационными сетями: учебное пособие. Новосибирск: СибГУТИ, 2001. 112 с.
- 2. *Stallings W.* SNMP, SNMPv2 and RMON: Practical Network Management. MA: Addison-Wesley, 1996.
- 3. Семенов В. Ю. Исследование и анализ средств и методов мониторинга вычислительных сетей. [Электронный pecypc]. URL: https://cyberleninka.ru/article/ n/issledovanie-i-analiz-sredstv-i-metodov-monitoringa-vychislitelnyhsetey/viewer (дата обращения: 05.03.2022).
- 4. Панов И. Часть 1. Пять ключевых функций систем мониторинга производительности сети [Электронный pecypc]. URL: https://networkguru.ru/piat-cliuchevykh-funktcii-sistem-monitoringa-proizvoditelnosti-seti/ (дата обращения: 07.03.2022).
- 5. Бродский Г. ТЕОСО представляет NETRAC 7, новое решение Service Assurance для операторов связи [Электронный ресурс]. URL: https://www.content-review.com/articles/21149/ (дата обращения: 05.03.2022).
- 6. *Рудаков П., Саякин В.* Управление производительностью сети [Электронный ресурс]. URL: https://www.bytemag.ru/articles/detail.php?ID=6609 (дата обращения: 10.03.2022).
- 7. *Евдаимов А.* Компания САТЕЛ. Аналитика больших данных «COBA». [Электронный реcypc]. URL: https://satel.org/solutions/programmnye-produkty/ (дата обращения: 06.03.2022).
- 8. Бряндинский А. Системы EMC SMARTS. [Электронный pecypc]. URL: https://comptek.ru/box/1404?printered=o (дата обращения: 07.03.2022).
- 9. Левашов A. Dell EMC. EMC Smarts IP Performance Manager [Электронный ресурс]. URL: https://www.tadviser.ru/a/188974 (дата обращения: 07.03.2022).

- 10. Какунин A. HP Network Node Manager. [Электронный ресурс]. URL: https://axoft.ru/vendors/Micro-Focus/HP-Network-Node-Manage/ (дата обращения: 04.03.2022).
- 11. *Корсунский А*. Обзор системы мониторинга Zabbix [Электронный ресурс]. URL: https://servergate.ru/articles/obzor-sistemy-monitoringa-zabbix/ (дата обращения: 08.03.2022).
- 12. Berry I. About Cacti. 2021. The Cacti Group, Inc.URL: https://www.cacti.net/ (дата обращения: 07.03.2022).
- 13. *Todeschini G*. The NOC Project. 2019. URL: https://www.getnoc.com/ (дата обращения: 08.03.2022).
- 14. Nocproject. Fault Managment система управления событиями и ошибками в NOC [Электронный pecypc]. URL: https://kb.nocproject.org/pages/viewpage.actio n?pageId=15106281 (дата обращения: 08.03.2022).
- 15. ПАО «Ростелеком». Система «Fault management/Performance management» на сегментах сетей доступа МРФ «Сибирь» на основе ПК «СОВА». Пояснительная записка. [Внутренний документ]. 2016.
- 16. ОАО «Сибирьтелеком». Система мониторинга сети широкополосного доступа. Технический проект. Пояснительная записка. [Внутренний документ]. 2009.
- 17. Владышев А. Исследуйте возможности Zabbix [Электронный ресурс] URL: https://www.zabbix.com/ru/features (дата обращения: 07.03.2022).
- 18. *Володин Д.* NOC: Комплексный подход к управлению сетью [Электронный pecypc] URL: https://habr.com/ru/post/125034/ (дата обращения: 06.03.2022).
- 19. Jain A. Performance Management Software [Электронный pecypc]. URL: https://www.teoco.com/products-services/service-assurance/performance-management/ (дата обращения: 06.03.2022).
- 20. ПАО «Ростелеком». Система мониторинга неисправностей и производительности IP/MPLS сети и система мониторинга неисправностей транспортной сети. Описание архитектуры решения. [Внутренний документ]. 2016.

Статья поступила в редакцию 15.03.2022; переработанный вариант – 05.04.2022.

Доценко Владислав Олегович

магистрант СибГУТИ, e-mail: dietaar98@gmail.com.

Шевнина Ирина Евгеньевна

к.т.н., доцент кафедры инфокоммуникационных систем и сетей СибГУТИ (630102, Новосибирск, ул. Кирова, 86), тел. (383) 269-83-44, e-mail: pdsm@yandex.ru.

Research of current requirements for the developments of telecommunications network monitoring and control systems

Vladislav O. Dotsenko

Master's student, Siberian State University of Telecommunications and Information Science (SibSUTIS, Novosibirsk, Russia), dietaar980gmail.com.

Irina E. Shevnina

Candidate of technical sciences, Docent, Siberian State University of Telecommunications and Information Science (SibSUTIS, Novosibirsk, Russia), pdsm@yandex.ru.

The modern telecommunications market dictates the need to modernize the existing requirements for monitoring and management systems of data transmission networks. The infrastructure of existing telecommunications networks is continuously becoming more complex and expanding, therefore, modern troubleshooting and network performance monitoring products must be equipped with the appropriate functionality. This paper determines the current requirements for modern systems, which must be taken into account when developing future monitoring and management software products. This article examines the functional features of software systems used on the networks of telecom operators, namely: TEOCO Netrac 7, SATEL SOVA, EMC Smarts, HP Network Node Manager, Zabbix, Cacti, NOC Project. Comparison and analysis are the tools used in this research. The research topic meets the criteria of the Ministry of Digital Development, Communications and Mass Communications of the Russian Federation (Federal Law No. 204-FL of June 23, 2016 "On Amendments to Articles 2 and 24 of the Federal Law "On Communications").

Keywords: monitoring and control systems, telecommunication network, troubleshooting.

References

- 1. Fokin V.G. *Upravlenie telekommunikacionnymi setjami* [Management of telecommunication networks]. Novosibirsk, SibGUTI, 2001, 112 p.
- 2. Stallings, W. SNMP, SNMPv2 and RMON: Practical Network Management. W. Stallings Readong, MA: Addison-Wesley, 199 p.
- 3. Semenov V.Ju. *Issledovanie i analiz sredstv i metodov monitoringa vychislitel'nyh setej* [Research and analysis of computer network monitoring tools and methods], available at: https://cyber-leninka.ru/article/n/issledovanie-i-analiz-sredstv-i-metodov-monitoringa-vychislitelnyh-setey/viewer (accessed 05.03.2022).
- 4. Panov. I. Chast' I. Pjat' kljuchevyh funkcij sistem monitoringa proizvoditel'nosti seti [Part 1. Five key functions of Network Performance Monitoring systems], 2015, available at: https://net-workguru.ru/piat-cliuchevykh-funktcii-sistem-monitoringa-proizvoditelnosti-seti/ (accessed 07.03.2022).
- 5. Brodskij G. *TEOCO predstavljaet NETRAC 7, novoe reshenie Service Assurance dlja operatorov svjazi* [TEOCO introduces NETRAC 7, a new Service Assurance solution for Telecom Operators], 2012, available at: https://www.content-review.com/articles/21149/ (accessed 05.03.2022).
- 6. Rudakov P., Sajakin V. *Upravlenie proizvoditel'nost'ju seti* [Network performance Management], 2002, available at: https://www.bytemag.ru/articles/detail.php?ID=6609 (accessed 10.03.2022).
- Evdaimov A. Kompanija SATEL. Analitika bol'shih dannyh «SOVA» [SATEL. Big data analytics "SOVA"], 2022, available at: https://satel.org/solutions/programmnye-produkty/ (accessed 06.03.2022).
- 8. Brjandinskij A. *Sistemy EMC SMARTS* [EMC SMARTS systems], 2022, available at: https://comptek.ru/box/1404?printered=o (accessed 07.03.2022).
- 9. Levashov A. *Dell EMC. EMC Smarts IP Performance Manager*, 2014, available at: https://www.tadviser.ru/a/188974 (accessed 07.03.2022).
- 10. Kakunin A. *HP Network Node Manager*, 2022, available at: https://axoft.ru/vendors/Micro-Focus/HP-Network-Node-Manage/ (accessed 04.03.2022).
- 11. Korsunskij A. *Obzor sistemy monitoringa Zabbix* [Overview of the Zabbix monitoring system], 2021, available at: https://servergate.ru/articles/obzor-sistemy-monitoringa-zabbix/ (accessed 08.03.2022).
- 12. Berry I. *About Cacti*, The Cacti Group, Inc., 2021, available at: https://www.cacti.net/ (accessed 07.03.2022).
- 13. Todeschini G. *The NOC Project*, 2019, available at: https://www.getnoc.com/ (accessed 08.03.2022).
- 14. Nocproject. *Fault Managment sistema upravlenija sobytijami i oshibkami v NOC* [Fault Managment is an event and error management system in the NOC. Blogs Confluence], 2015, available at: https://kb.nocproject.org/pages/viewpage.action?pageId=15106281 (accessed 08.03.2022).

- 15. Zenin N. Sistema «Fault management/Performance management» na segmentah setej dostupa MRF «Sibir» na osnove PK «SOVA» [The "Fault management/Performance management" system on the segments of the access networks of the MRB "Siberia" based on the program complex "SOVA"]. Saint Petersburg, Rostelekom, 2016.
- 16. Dadykin I. Sistema monitoringa seti shirokopolosnogo dostupa. Tehnicheskij proekt [Broadband access network monitoring system. Technical project]. Saint Petersburg, Sibirtelekom, 2009.
- 17. Vladyshev A. Zabbix features overview, 2022, available at: https://www.zabbix.com/ru/features (accessed 07.03.2022).
- 18. Volodin D. *NOC: Kompleksnyj podhod k upravleniju set'ju* [NOC: An integrated approach to network management], 2011, available at: https://habr.com/ru/post/125034/ (accessed 06.03.2022).
- 19. Jain A. *Performance Management Software*, 2022, available at: https://www.teoco.com/prod-ucts-services/service-assurance/performance-management/ (accessed 06.03.2022).
- 20. Oseevsky M. Sistema monitoringa neispravnostej i proizvoditel'nosti IP/MPLS seti i sistema monitoringa neispravnostej transportnoj seti. Opisanie arhitektury reshenija. [Fault monitoring system and IP/MPLS network performance and transport network fault monitoring system. Description of the solution architecture]. Saint Petersburg, Rostelekom, 2016.

УДК 004.057.4 DOI: 10.55648/1998-6920-2022-16-2-33-39

Криптографический протокол поиска места встречи участников со свойством конфиденциальности

И. Д. Иогансон, А. А. Голованов, Ж.-М. Н. Дакуо, В. В. Давыдов¹

Предложен новый криптографический протокол поиска места встречи участников. Такой протокол позволяет нескольким участникам выбрать место встречи, наиболее близкое к их местоположениям, при этом не раскрывая координаты каждого из участников. Протокол также позволяет детектировать ошибку, которая возникла в ходе передачи информации между участниками. Доказана корректность и безопасность используемого подхода, определены его основные достоинства и недостатки.

Ключевые слова: криптографический протокол, свойство конфиденциальности, протоколы конфиденциального вычисления.

1. Введение

Криптографические протоколы занимают очень важное место в современных информационных системах. Построение надёжных протоколов определяет безопасность системы и её корректное функционирование.

Одними из важнейших на сегодняшний день криптографических схем являются протоколы конфиденциального вычисления. Основная идея таких протоколов заключается в том, чтобы вычислить результат некоторой функции от закрытых данных нескольких участников, при этом не нарушая конфиденциальности.

История конфиденциальных вычислений началась с работы Эндрю Яо о «задаче двух миллионеров», в которой два участника хотят выяснить, кто из них богаче, не раскрывая размеры своих состояний [1]. Яо предложил защищённый протокол, позволяющий решить эту задачу, а также ввёл понятие протоколов конфиденциального вычисления. В настоящее время подобные протоколы находят широкое применение в различных областях. Например, они широко используются в медицине для обучения моделей диагностики без раскрытия данных пациентов [2]. Также распространено использование в статистике, например, для анализа различия в заработной плате в зависимости от пола без раскрытия точного размера зарплаты [3].

В данной работе уделено внимание протоколам, которые позволяют вычислить сумму секретных значений нескольких пользователей. В работе [4] приведён пример простейшего варианта такого протокола. Он имеет следующую структуру:

1. Пользователям присваиваются номера от 1 до k.

2. Пользователь 1 генерирует случайное число *m*₀.

3. Далее, начиная с пользователя 1, каждый пользователь i получает значение m_{i-1} от пользователя i-1 (или в случае пользователя 1 это будет просто сгенерированное им значение

¹ Работа выполнена при поддержке программы «Приоритет 2030».

 m_0) и отправляет пользователю i+1 (пользователь k отправит свое значение пользователю 1) значение $m_i = m_{i-1} + x_i$, где x_i – входное значение пользователя i.

4. В конце пользователь 1 вычисляет значение $m_k - m_0$, равное искомой сумме, и отправляет его остальным пользователям.

В статье [5] описан так называемый «k-Secure Sum Protocol». Этот протокол для *n* пользователей выглядит следующим образом:

- 1. Пусть *k* целое число, обозначающие параметр безопасности.
- 2. Пусть пользователи пронумерованы от 0 до n-1.
- 3. Пусть D_i входное значение, принадлежащее пользователю *i*.
- 4. Разобъём D_i на k сегментов $D_{i0}, ..., D_{i(k-1)}$ таким образом, что $D_i = \sum_{i=0}^{k-1} D_{ij}$.
- 5. Пусть $S_{00} = 0$.
- 6. Для *j* от 0 до *k*-1:
 - а. Для *i* от 0 до *n*-1:
 - і. Пользователь *i* посылает пользователю $(i+1) \mod n$: $S_{(i+1)j} = S_{ij} + D_{ij}$. b. $S_{0(i+1)} = S_{ni}$.

b.
$$S_{0(j+1)} = S_{nj}$$
.

7. Пользователь 0 публикует значение S_{0k} .

У приведённых выше протоколов есть схожий недостаток: в самом конце первый пользователь получает результат и публикует его. Проблема заключается в том, что если первый пользователь является злоумышленником, то ему не составит труда подменить результат на нужный ему.

В данной работе предлагается новый протокол для конфиденциального вычисления места встречи группы людей, которые хотят сохранить своё исходное местоположение в тайне. Пусть k пользователей P_1, P_2, \ldots, P_k хотят договориться о встрече. Исходное местоположение каждого пользователя выражено числом $x_i \in \mathbb{R}$, где $i \in \{1...k\}$, которое является координатой в некоторой системе координат. Требуется найти координату встречи y, наиболее удобную для всех участников, что в данной работе понимается как среднее арифметическое значение:

$$y = \frac{\sum_{i=1}^{k} x_i}{k} \,. \tag{1}$$

При этом предлагаемый протокол обладает следующими свойствами:

1. Значение x_i не раскрывается пользователем P_i ни на одном из этапов.

2. Нет необходимости доверять единственному пользователю вычисление и публикацию итогового значения, так как каждый пользователь может вычислить *у*.

3. В генерации случайных чисел участвуют все пользователи. Это необходимо, чтобы не концентрировать ответственность за случайную компоненту на единственном пользователе. Если за генерацию случайных чисел отвечает только один пользователь, то он может использовать нестойкий алгоритм генерации псевдослучайных чисел, который позволит предсказывать получаемые значения. Это может открыть возможности для атак на конфиденциальность системы.

2. Описание протокола

Протокол разбит на несколько этапов: инициализация, первый – третий круги и завершение. Опишем каждый этап более подробно.

1. Инициализация.

На этапе инициализации пользователи формируют защищенные каналы между соседями так, как это продемонстрированно на рис. 1. По необходимости каждый пользователь представляет свое значение x_i в виде целого числа умножением на 10^r , где r – это максимально допустимое количество значащих цифр после запятой. Далее каждый пользователь формирует обязательство, подобное тому, что использовал П. Фельдман в своей проверяемой схеме разделения секрета [6], для своего значения $x_i: C_i = x_i \cdot G$, где G – генераторная точка на эллиптической кривой (ЭК) E над полем F_q , где q – простое. ЭК E и точка G известны всем пользователям. Обязательства $C_1...C_k$ публикуются в свободном доступе для каждого пользователя.



Рис. 1. Общий вид схемы

2. Первый круг.

Каждый пользователь выбирает случайное значение $s_i \in_R \mathbb{Z}$. Далее начинается передача сообщений. Обозначение $P_a \to P_b$: обозначает, что пользователь P_a передаёт информацию пользователю P_b , после двоеточия описана передаваемая информация. Также будем считать, что $P_{k+1} = P_1$.

Передаваемая информация вычисляется по правилу:

$$P_i \to P_{i+1} : m_{1,i} = m_{1,i-1} + x_i + s_i,$$
(2)

где $m_{1,0} = 0$.

3. Второй круг.

Пусть

$$m_{2,0} = m_{1,k} = \sum_{i=1}^{k} (x_i + s_i).$$
(3)

Тогда правила вычисления передаваемой информации на данном круге:

$$P_i \to P_{i+1} : m_{2,i} = m_{2,i-1} - 2s_i - x_i.$$
(4)

4. Третий круг.

Пусть

$$m_{3,0} = m_{2,k} = -\sum_{i=1}^{k} s_i.$$
⁽⁵⁾

Тогда правила вычисления передаваемой информации на данном круге:

$$P_i \to P_{i+1} : m_{3,i} = m_{3,i-1} + s_i.$$
(6)

Если в конце протокола пользователь P_1 получил значение $m_{3,k}$. Если $m_{3,k} = 0$, то протокол считается корректно выполненным, и P_1 рассылает сообщение о корректном окончании протокола. Иначе – во время работы протокола возникла ошибка, и P_1 рассылает сообщение о том, что протокол окончился с ошибкой.

5. Завершение.

Каждый пользователь P_i обладает набором значений $\{m_{1,i}, m_{2,i}, m_{3,i}\}$, полученных им на кругах 1, 2 и 3 соответственно. Далее он вычисляет:

$$y = \frac{m_{1,i} + m_{2,i} + m_{3,i}}{k} \,. \tag{7}$$

Данное значение и является координатой сбора.

Также каждый пользователь проводит валидацию полученного значения с помощью обязательств $C_1 \dots C_k$, полученных на этапе инициализации, по формуле:

$$\sum_{i=1}^{k} C_i = (k \cdot y) \cdot G.$$
(8)

3. Доказательства корректности и безопасности

3.1. Доказательство корректности

На конце 1-го и 2-го кругов пользователь P_1 получает значения:

$$m_{1,k} = \sum_{j=1}^{k} (x_j + s_j);$$
(9)

$$m_{2,k} = -\sum_{j=1}^{k} s_j.$$
(10)

Пользователь P_i по итогу каждого круга имеет значения:

$$m_{1,i} = \sum_{j=1}^{i} (x_j + s_j);$$
(11)

$$m_{2,i} = \sum_{j=1}^{k} (x_j + s_j) - \sum_{j=1}^{i} (x_j + 2s_j) = \sum_{j=i+1}^{k} (x_j + s_j) - \sum_{j=1}^{i} s_j;$$
(12)

$$m_3 = -\sum_{j=1}^k s_j + \sum_{j=1}^i s_j = -\sum_{j=i+1}^k s_j.$$
(13)
Тогда

$$y = \frac{1}{k} (m_{1,i} + m_{2,i} + m_{3,i}) =$$

$$= \frac{1}{k} (\sum_{j=1}^{i} (x_j + s_j) + \sum_{j=i+1}^{k} (x_j + s_j) - \sum_{j=1}^{i} s_j - \sum_{j=i+1}^{k} s_j) = \frac{1}{k} \sum_{j=1}^{k} x_j.$$
(14)

3.2. Доказательство безопасности

Каждый пользователь *i* знает следующий набор значений: $\{k, x_i, s_i, m_{1i}, m_{2i}, m_{3i}, y\}$. Обозначим:

$$x_{a} = \sum_{j=1}^{i-1} x_{j}, \ x_{b} = \sum_{j=i}^{k} x_{j};$$

$$s_{a} = \sum_{j=1}^{i-1} s_{j}, \ s_{b} = \sum_{j=i}^{k} s_{j}.$$
(15)

Тогда сообщения m_{1i}, m_{2i}, m_{3i} :

$$m_{1i} = x_a + s_a;$$

 $m_{2i} = x_b + s_b - s_a;$
 $m_{3i} = -s_b.$
(16)

Система из четырёх уравнений с четырьмя переменными x_a, x_b, s_a, s_b :

$$\begin{cases} x_{a} + s_{a} = m_{1i} \\ x_{b} + s_{b} - s_{a} = m_{2i} \\ -s_{b} = m_{3i} \\ x_{a} + x_{b} = y^{*}k \end{cases}$$
(17)

является линейно зависимой, так как сумма первых трех уравнений равна четвертому:

$$+\begin{cases} x_a + s_a = m_{1i} \\ x_b + s_b - s_a = m_{2i} \\ -s_b = m_{3i} \end{cases}$$
(18)

$$x_a + x_b = m_{1i} + m_{2i} + m_{3i} = y * k$$

Следовательно, данная система не имеет единого решения относительно этих переменных. Таким образом, из имеющегося набора значений пользователь i не может получить информацию об x_i других пользователей (кроме y).

Применяемая схема обязательств основана на проблеме дискретного логарифмирования [7] и не позволяет злоумышленнику раскрыть значение секрета. Также в представленной схеме обязательств отсутствует возможность подмены значения секрета создателем этого обязательства.

4. Оценки протокола

4.1. Обобщение

Данный протокол в базовом варианте работает в 1-мерном пространстве, однако его очевидным образом можно обобщить на *n*-мерное пространство, где координаты каждого пользователя *i* задаются *n*-мерным вектором $X_i = (x_{i1}, x_{i2}, ..., x_{in})$. В таком случае необходимо просто произвести данный протокол для каждой из координат по отдельности и на выходе получить $Y = (y_1, y_2, ..., y_n)$.

4.2. Ограничения

Данный протокол проводится для k участников. Однако при k = 2 оба пользователя очевидным образом, зная $y = \frac{x_1 + x_2}{2}$ и свой x_i , могут найти x_{3-i} другого пользователя. Поэтому число k участников протокола всегда должно быть больше 2.

Также в случае, если у пользователя *i* были скомпрометированы оба ключа $K_{(i-1)i}$ и $K_{i(i+1)}$, то злоумышленник, перехватив сообщения, которые пользователь принимает и посылает в процессе действия протокола, сможет однозначно установить значение x_i . Аналогичное действие могут совершить пользователи i+1 и i-1, если сговорятся и поделятся друг с другом сообщениями, которые они посылали и принимали от пользователя *i*.

Литература

- 1. *Yao A. C.* Protocols for secure computations // 23rd IEEE Annual symposium on foundations of computer science (SFCS), 1982. P. 160–164.
- 2. *Dugan T., Zou X.* A survey of secure multiparty computation protocols for privacy preserving genetic tests // 2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), 2016. P. 173–182.
- 3. *Lapets A. et al.* Secure MPC for analytics as a web application // 2016 IEEE Cybersecurity Development (SecDev), 2016. P. 73–74.
- 4. *Clifton C. et al.* Tools for privacy preserving distributed data mining // ACM Sigkdd Explorations Newsletter. 2002. V. 4, №. 2. P. 28–34.
- 5. *Sheikh R., Kumar B., Mishra D. K.* Privacy preserving k secure sum protocol // arXiv preprint arXiv:0912.0956. 2009.
- 6. *Feldman P*. A practical scheme for non-interactive verifiable secret sharing // 28th IEEE Annual Symposium on Foundations of Computer Science (SFCS), 1987. P. 427–438.
- 7. *Silverman J. H., Pipher J., Hoffstein J.* An introduction to mathematical cryptography. Springer New York, 2008.

Статья поступила в редакцию 25.04.2022; переработанный вариант – 01.05.2022.

Иогансон Иван Дмитриевич

инженер факультета безопасности информационных технологий Университета ИТМО (197101, Санкт-Петербург, Кронверкский просп., д. 49, литер А), e-mail: ivan.ioganson@yandex.ru.

Голованов Андрей Андреевич

инженер факультета безопасности информационных технологий Университета ИТМО, e-mail: agolovanov2403@gmail.com.

Дакуо Жан-Мишель Никодэмович

инженер факультета безопасности информационных технологий Университета ИТМО, e-mail: jeandakuo@mail.ru.

Давыдов Вадим Валерьевич

преподаватель факультета безопасности информационных технологий Университета ИТМО, e-mail: vvdavydov@itmo.ru.

Cryptographic protocol for finding the meeting place of participants with confidentiality property

Ivan D. Ioganson

Engineer, ITMO University (St. Petersburg, Russia), ivan.ioganson@yandex.ru.

Andrei A. Golovanov

Engineer, ITMO University (St. Petersburg, Russia), agolovanov2403@gmail.com.

Zhan-Mishel N. Dakuo

Engineer, ITMO University (St. Petersburg, Russia), jeandakuo@mail.ru.

Vadim V. Davydov

Lecturer, ITMO University (St. Petersburg, Russia), vvdavydov@itmo.ru.

In this paper a new cryptographic protocol for finding the meeting place of participants is proposed. This protocol allows several participants to choose the meeting place closest to their locations not revealing the coordinates of each of the participants. The protocol also allows to detect an error occurring during the transfer of information between participants. The correctness and security of the used approach are proved, and the main advantages and disadvantages are determined.

Keywords: cryptographic protocol, confidentiality property, secure multi-party computation.

References

- 1. Yao A. C. Protocols for secure computations. 23rd Annual Symposium on Foundations of Computer Science (sfcs 1982), 1982, pp. 160-164. DOI: 10.1109/SFCS.1982.38.
- 2. Dugan T., Zou X. A survey of secure multiparty computation protocols for privacy preserving genetic tests. 2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), 2016, pp. 173-182. DOI: 10.1109/CHASE.2016.71.
- 3. Lapets A., Volgushev N., Bestavros A., Jansen F. and Varia M. Secure MPC for analytics as a web application. 2016 IEEE Cybersecurity Development (SecDev), 2016, pp. 73-74. DOI: 10.1109/SecDev.2016.027.
- 4. Clifton C., Kantarcioglu M., Jaideep V., Lin X. and Zhu M. Y. Tools for privacy preserving distributed data mining. *ACM SIGKDD Explorations*, 2003, vol. 4, pp. 28-34.
- 5. Sheikh R., Kumar B., Mishra D. K. Privacy preserving k secure sum protocol. *International Journal of Computer Science and Information Security*, 2009, vol. 6.
- 6. Feldman P. A practical scheme for non-interactive verifiable secret sharing. 28th Annual Symposium on Foundations of Computer Science (sfcs 1987), 1987, pp. 427-438. DOI: 10.1109/SFCS.1987.4.
- 7. Silverman J. H., Pipher J., Hoffstein J. *An introduction to mathematical cryptography*, Springer New York, 2008. vol 1.

Определение оптимальной длины кадра в сети наземного цифрового телевизионного вещания

В. И. Носов, А. С. Ладан, М. В. Зиновьев

В работе предлагается простая методика определения оптимальной длины кадра в стандарте DVB-T2, основанная на простых и понятных формулах, в основу которых положена последовательность обработки пользовательских данных в стандарте DVB-T2. В T2-кадре всегда должно быть целое число FEC-блоков, оставшиеся ячейки, которые не составляют целый FEC-блок, как правило, будут фиктивными ячейками и представляют собой потерю скорости. В результате общая скорость передачи битов и глубина перемежения в мельчайших деталях зависят от точного количества OFDM-символов, а следовательно, и ячеек в T2-кадре.

Ключевые слова: стандарт DVB-T2, T2-кадр, OFDM-символ, FEC-блок, фиктивные ячейки, глубина перемежения, оптимальная длина кадра.

1. Введение

Несмотря на то, что увеличение длины кадра уменьшает накладные расходы преамбулы, существует эффект второго порядка, который должен быть учтён при выборе оптимального количества OFDM-символов в T2-кадре. OFDM-символ состоит из ячеек, количество бит в каждой из которых определяется позиционностью модуляции. Количество ячеек в одном OFDM-символе определяется числом используемых поднесущих частот (Subcarrier Frequency, SCF). В T2-кадре число FEC-блоков всегда должно быть целым, оставшиеся ячейки, которые не составляют целый FEC-блок, как правило, будут фиктивными (пустыми, dummy) ячейками и представляют собой потерю емкости [1, 5]. В результате общая скорость передачи битов и глубина перемежения в мельчайших деталях зависят от точного количества символов, а следовательно, и ячеек в кадре. Изменение длины кадра на один символ часто может привести к изменению глубины перемежения и общей скорости передачи битов за счет уменьшения количества фиктивных ячеек. Увеличение глубины перемежения OFDM-символов позволяет увеличить помехоустойчивость приёма при наличии селективных замираний сигнала в канале при распространении радиоволн. Для решения поставленной задачи определения оптимального количества OFDM-символов в T2-кадре необходимо детально рассмотреть работу блока временного перемежения (Time Interliving, TI).

В данной работе предлагается простая методика определения оптимальной длины кадра с точки зрения глубины перемежения и скорости передачи в стандарте DVB-T2, основанная на простых и понятных формулах, в основу которых положена последовательность обработки пользовательских данных в стандарте DVB-T2.

2. Определение глубины перемежения символов в кадре Т2

Вход в систему T2 (рис. 1) осуществляется одним или несколькими транспортными потоками данных TS (Transport Stream, TS). Транспортный поток TS – это последовательность пользовательских пакетов (User Packet, UP) фиксированной длины: $UP_L = 188 \times 8$ бит (один MPEG-пакет).



Рис. 1. Обобщенная схема обработки передаваемых сигналов в системе DVB-T2

После обработки сигнала в модуле адаптации («Входная обработка» на рис. 1) сигналы потоков поступают на обработку в модуль перемежения бит, кодирования и модуляции (Bit Interleaving Coding Modulation, BICM) (рис. 1). На рис. 2 представлена схема обработки одного потока в блоке BICM.



Рис. 2. Модуль с перемежением, кодированием и модуляцией (BICM)

Первый блок в модуле BICM – кодер FEC, где добавляются избыточные биты, которые используются в декодере FEC для обнаружения и исправления ошибочных бит. В результате на выходе кодера формируется кадр FEC (FEC_FRAME). В DVB-T2 системе используется каскадное кодирование, где в качестве внешнего кода используется блочный код BCH, а внутреннего – низкоплотностный код LDPC (рис. 2, 3).



Рис. 3. Формат данных на выходе кодера FEC

Из последовательности обработки пользовательских данных в стандарте DVB-T2 (рис. 2, 3) следует, что скорость внешнего кода ВСН зависит от скорости внутреннего кода LDPC R_{c_LDPC}, на выходе которого всегда получается нормальный блок FEC (BLOCK_FEC) одина-ковой длительности [1, 5]:

$$N_b$$
 BLOCK FEC = N_b LDPC = 64800 бит. (1)

Количество ячеек (cells) в блоке FEC зависит от позиционности модуляции M_{mod}, используемой в стандарте DVB-T2:

$$N_{CELL_IN_BLOCK_FEC} = \frac{N_{b_BLOCK_FEC}}{\log_2 M_{mod}}.$$
 (2)

Для решения поставленной задачи определения оптимального количества OFDMсимволов в T2-кадре необходимо детально рассмотреть работу блока временного перемежения. Блоки FEC с выхода перемежителя ячеек должны быть сгруппированы в чередующиеся TI-блоки, которые отображаются на один T2-кадр (рис. 4). Каждый кадр перемежения (Interliving Frame, IF) должен содержать динамически изменяемое целое число блоков FEC. Количество блоков FEC N_{BLOCK_FEC_IN_TI_BLOCK} в TI-блоке временного перемежения может варьироваться от минимального значения 0 до максимального значения NMAX BLOCK FEC_IN_TI_BLOCK.

Каждый кадр перемежения отображается непосредственно на один T2-кадр. Каждый кадр перемежения также разделен на один или несколько (N_{TI}) TI-блоков, где TI-блок соответствует одному использованию памяти временного перемежителя. TI-блоки внутри кадра перемежения могут содержать разное количество блоков FEC. Если кадр перемежения разделен на несколько TI-блоков, он должен быть сопоставлен только одному T2-кадру (рис. 4).



Рис. 4. Структура временного перемежения в Т2-кадре

Блок временного перемежения TI (рис. 5) должен быть перемежителем блоков строк и столбцов. Число строк N_r в TI-блоке перемежения равно числу ячеек в блоке FEC (Ncells in Block Fec), деленному на 5:

$$N_{row_IN_TI} = \frac{N_{CELL_IN_BLOCK_FEC}}{5}.$$
(3)

Количество строк в ТІ-блоке перемежения зависит от позиционности модуляции (табл. 1).

Длина блока FEC, бит	Позиционность модуляции	Количество ячеек в FEC-блоке	Количество строк в блоке временного перемежения
	256-QAM	8100	1620
64800	64-QAM	10800	2160
	16-QAM	16200	3240
	QPSK	32400	6480

Таблица 1. Параметры для временного перемежителя

Графическое представление временного перемежителя показано на рис. 5. Первый блок FEC записывается по столбцам в первые 5 столбцов временного перемежителя, второй блок FEC записывается по столбцам в следующие 5 столбцов и так далее. Ячейки считываются по строкам.



Рис. 5. Блок временного перемежения TI

Число столбцов в TI-блоке перемежения N_{COLLS} определяется количеством блоков FEC в TI-блоке перемежения N_{BLOCK FEC IN TI}:

$$N_{COLLS \ IN \ TI} = 5 \times N_{BLOCK \ FEC \ IN \ TI} .$$
(4)

Количество блоков FEC в TI-блоке перемежения зависит от ёмкости блока и количества OFDM-символов в T2-кадре. Максимальное количество ячеек памяти в одном блоке временного перемежителя [1]:

$$N_{CELL_MAX_IN_TI} = 2^{19} + 2^{15} = 557056 \text{ ячеек.}$$
(5)

Максимальное количество блоков FEC, которое можно разместить в одном блоке временного перемежителя равно:

$$N_{BLOCK_FEC_in_TI} = \frac{N_{CELL_MAX_IN_TI}}{C_{BLOCK_FEC}}.$$
(6)

Сформированный в модуле перемежения, кодирования и модуляции (BICM) цифровой сигнал поступает на формирователь кадра (Frame Builder) (рис. 1). Самый большой объект системы DVB-T2 – это суперкадр, который состоит из T2-кадров (рис. 6).



Рис. 6. Структура кадров в системе DVB-T2

Максимальное число T2-кадров в суперкадре равно 255. Максимальная длина T2-кадра Т_{FRAME_T2_MAX} составляет:

$$T_{FRAME T2 MAX} = 250 \,\mathrm{mc} \,. \tag{7}$$

Однако все Т2-кадры имеют одинаковую длину внутри суперкадра. Т2-кадр состоит из символов OFDM, каждый из которых имеет активную длительность T_U. Т2-кадр состоит из: преамбулы, состоящей из одного символа P1; второй преамбулы, состоящей из нескольких символов P2; настраиваемого количества символов данных PLP_i (рис. 7).



Рис. 7. Распределение фрагментов потоков в Т2-кадре

Длительность T2-кадра определяется размером ОБПФ, защитным интервалом и числом используемых OFDM-символов. Максимально допустимая длительность T2-кадра составляет 250 мс, что накладывает ограничение на максимальное количество символов OFDM NOFDM_SYMB для различных размеров БПФ и защитных интервалов (табл. 2). Длительность T2-кадра T_{T2} FRAME вычисляется по формуле:

$$T_{T2} \quad FRAME = N_{OFDM} \quad SYMB \times T_S + T_{P1}, \tag{8}$$

где Т_{ОFDM SYMB} – общая длина OFDM-символа, определяемая выражением:

$$T_{OFDM} \quad SYMB = T_S = T_U + T_{GI} = T_U \times (1 + GIF).$$

$$\tag{9}$$

В (14) GIF – доля защитного интервала T_{GI} относительно активной длительности OFDM-символа T_{U} :

$$GIF = T_{GI} / T_{U}. \tag{10}$$

Длительность символа P1 составляет:

$$T_{P1} = 0.224 \,\mathrm{Mc} \,. \tag{11}$$

Количество OFDM-символов в кадре T2 N_{OFDM_SYMB} включает все P2-символы N_{P2} и символы данных $N_{DATA\ SYMB}$:

$$N_{OFDM} \quad SYMB = N_{P2} + N_{DATA} \quad SYMB .$$
⁽¹²⁾

При этом общее количество ячеек С_{ТОТАL Т2 FRAME} в кадре Т2 будет равно:

$$C_{TOTAL_T2_FRAME} = N_{OFDM_SYMB} \cdot N_{SCF_USED}.$$
⁽¹³⁾

Кадр физического уровня DVB-T2 (T2-кадр) (рис. 7) начинается с преамбулы P1. Этот ОFDM-символ служит для синхронизации, идентификации потока DVB-T2, а также содержит семь информационных бит с начальной информацией о T2-кадре, а именно: число номинальных поднесущих частот в OFDM-символе (1К–32К) и формат передачи следующей за P1 преамбулы P2. Вся остальная информация о T2-кадре (длина данных, модуляция, скорость кодирования и т.п.) передается в преамбуле P2, которая может занимать несколько OFDM-символов. Далее следует поле данных (информационные OFDM-символы).

Максимальное число OFDM-символов N_{OFDM_SYMB} в кадре T2 можно определить из уравнений (14) – (19), учитывая, что T_{T2} _{FRAME MAX} = 250 мс:

$$N_{ODM} _ SYMB _ MAX = \frac{T_{T2} _ FRAME _ MAX - T_{P1}}{T_{OFDM} _ SYMB} = \frac{T_{T2} _ FRAME _ MAX - T_{P1}}{T_U \times (1 + GIF)}.$$
 (14)

Рассчитанное по выражению (14) максимальное число OFDM-символов N_{OFDM_SYMB} в кадре Т2 приведено в табл. 2. В зависимости от параметров OFDM-символа в T2-кадре может быть от 60 до 2098 OFDM-символов при полосе передачи 8 МГц.

Таблица 2. Максимальное число OFDM-символов $N_{OFDM_SYMB_MAX}$ в кадре T2 для различных значений IFFT size и GIF для полосы канала 8 МГц, включая символы N_{P2} и символы данных

IFFT	Τ (GIF					
size	$I_{\rm U}$ (ms)	1/128	1/32	1/16	19/256	1/8	19/128	1/4
32 K	3.584	68	66	64	64	60	60	н.д.
16 K	1.792	138	135	131	129	123	121	111
8 K	0.896	276	270	262	259	247	242	223
4 K	0.448	н.д.	540	524	51219	495	485	446
2 K	0.224	н.д.	1081	1049	1038	991	970	892
1 K	0.112	н.д.	н.д.	2098	2076	1982	1941	1784

N_{DATA_SYMB}

Количество символов Р2 N_{P2} [2] зависит от размера ОБПФ и приведено в табл. 3.

IFFT Size	N _{P2}
1 K	16
2 K	8
4 K	4
8 K	2
16 K	1
32 K	1

Таблица 3. Зависимость числа символов Р2 N_{P2} от размера БПФ

Исходя из проведённого анализа обработки сигнала в блоке формирования кадров T2, можно считать, что при оптимальном выборе количества символов OFDM в кадре T2 единственными символами OFDM, содержащими как служебную, так и пользовательскую информацию, являются символы P2 (табл. 4) [1].

IFFT Size	C _{P2}
1 K	558
2 K	1118
4 K	2236
8 K	4472
16 K	8944
32 K	22432

Таблица 4. Количество ячеек пользовательских данных CP2 в одном символе P2

В DVB-T2 формируются 8 последовательностей пилот-ячеек PP1, PP2, ..., PP8 (табл. 5), которые могут быть выбраны в зависимости от размера БПФ и доли защитного интервала, принятых для конкретной передачи.

С учётом используемой последовательности пилот-ячеек изменяется количество доступных ячеек данных С_{data} в одном OFDM-символе (табл. 5).

Таблица 5. Количество доступных ячеек данных Сdata в одном OFDM-символе

		C _{data} , cells							
FFT size		PP1	PP2	PP3	PP4	PP5	PP6	PP7	PP8
1K		764	768	798	804	818			
2K		1522	1532	1596	1602	1632		1646	
4K		3084	3092	3228	3234	3298		3328	
8K	Norm	6208	6298	6584	6588	6728		6698	6698
	Extend	6296	6298	6584	6588	6728		6788	6788
16K	Norm	12418	12436	12988	13002	13272	13288	13416	13406
	Extend	12678	12698	13262	13276	13552	13568	13698	13688
32K	Norm		24886		26022		26592	26836	26812
	Extend		25412		26572		27152	27404	27376

При расчёте нужно учесть фиктивные ячейки, которые появятся, если в Т2-кадре количество блоков временного перемежения будет не целое.

Количество OFDM-символов данных L_F (длина кадра) в T2-кадре:

$$N_{OFDM} __{SYMB} = L_F = N_{P2} + L_{data} .$$
⁽¹⁵⁾

Из (15) следует, что количество символов данных равно:

$$L_{data} = L_F - N_{P2}. \tag{16}$$

Общее количество ячеек данных в Т2-кадре:

$$C_{total \ T2 \ F} = N_{P2} \cdot C_{P2} + L_{data} \cdot C_{data} \,. \tag{17}$$

Количество ячеек $C_{P2} = \theta(FFT_{size})$ и $C_{data} = \theta(FFT_{size}, PP)$ берутся из табл. 9 и 10 [1], причём $C_{data} - 6es$ учёта пилот-ячеек. Здесь $FFT_{size} - размер преобразования Фурье, PP – пилот-ячейки.$

Подставляя (6) в (4), определим количество FEC-блоков в кадре T2:

$$N_{BLOCK_FEC_in_T2} = \frac{N_{P2} \cdot C_{P2} + L_{data} \cdot C_{data}}{C_{BLOCK_FEC}}.$$
(18)

Максимальное количество блоков FEC, которое можно разместить в одном блоке временного перемежителя, с учётом выражения (5) равно

$$N_{MAX_BLOCK_FEC_in_TI} = \frac{N_{C_MAX_TI}}{C_{BLOCK_FEC}}.$$
(19)

Из выражений (18) и (19) определим максимальное количество блоков временного перемежения N_{TI_MAX} в кадре T2:

$$N_{TI_MAX} = \frac{N_{BLOCK_FEC_in_T2}}{N_{MAX_BLOCK_FEC_in_TI}}.$$
(20)

Длительность кадра T2 определяется произведением количества OFDM-символов $N_{OFDM SYMB} = L_F$ и длительности одного OFDM-символа:

$$T_{FRAME_T2} = N_{OFDM_SYMB} \times T_{OFDM_SYMB} .$$
⁽²¹⁾

Длительность одного OFDM-символа определяется в (9), при этом доля активной части символа T_U составит:

$$T_{U_FRACT} = \frac{T_U}{T_S} = \frac{T_U}{T_U + T_{GI}}.$$
(22)

Время, отводимое на один блок временного перемежения (глубина перемежения), будет равно:

$$T_{TI} = \frac{T_{FRAME} T^2}{N_{TI}}.$$
(23)

Количество блоков FEC в одном блоке временного перемежения равно:

$$N_{BLOCK_FEC_IN_TI} = \frac{N_{BLOCK_FEC_MAX_IN_TI} \times N_{OFDM_SYMB_IN_TI}}{N_{OFDM_SYMB_MAX_IN_TI}}.$$
 (24)

Глубина перемежения в символах с учётом (10) и (11) будет равна:

$$N_{SYMB_int_depth} = \frac{T_{TI}}{T_{OFDM_SYMB}} = \frac{T_{FRAME_T2}}{N_{TI} \left[T_U \times (1 + GIF) \right]}.$$
(25)

При этом количество блоков временного перемежения N_{TI} в кадре T2 с заданным числом OFDM-символов L_F в нём можно определить с использованием выражения:

$$N_{TI_in_T2F} = \frac{N_{OFDM_SYMB_in_T2F}}{N_{MAX_OFDM_SYMB_in_TI}} = \frac{L_F}{N_{MAX_OFDM_SYMB_in_TI}}.$$
 (26)

В (26) максимальное количество OFDM-символов в блоке временного перемежения можно определить по выражению:

$$N_{MAX}_OFDM_SYMB_in_TI = \frac{N_C_MAX_TI}{N_C_in_OFDM_SYMB} = \frac{N_C_MAX_TI}{N_{SCF}_in_OFDM_SYMB}.$$
 (27)

Из (24) определим количество фиктивных ячеек, используя десятичные знаки количества блоков FEC в T2-кадре:

$$C_{dummy} = \left(N_{BLOCK_FEC_in_T2F} - N_{BLOCK_FEC_IN_TI} \times N_{TI_in_T2F}\right) \cdot C_{BLOCKFEC} .$$
(28)

При этом количество ячеек данных в Т2-кадре (17) с учётом фиктивных ячеек составит:

$$C_{T2F} = N_{P2} \cdot C_{P2} + L_{data} \cdot C_{data} - C_{dummy}.$$
⁽²⁹⁾

Уменьшение скорости передачи пользовательских данных из-за использования фиктивных ячеек составит:

$$\Delta R_{bUP} = \frac{C_{T2F}}{C_{totalT2F}} = \frac{N_{P2} \cdot C_{P2} + L_{data} \cdot C_{data} - C_{dummy}}{N_{P2} \cdot C_{P2} + L_{data} \cdot C_{data}} \,. \tag{30}$$

Скорость передачи пользовательской информации R_{b_UP} [5] уменьшается относительно суммарной скорости передачи $R_{b\Sigma}$ из-за необходимости: передавать служебную информацию в сигнале основной полосы OHF_{BB_FRAME}, в кадре T2 N_{P2}, в пилот-сигналах OHF_{PP}; вводить избыточное кодирование FEC $R_{cFEC} = R_{cBCH} \times R_{cLDPC}$ и защитный интервал T_{U_FRACT} :

$$R_{b_UP} = \left[\frac{\log_2 M_{MOD}}{\left[N_{OFDM_{SYMB}} \times T_U \times (1 + GIF) + T_{P1} \right]} \times \left(1 - OHF_{BB_FRAME} \right) \times R_{c_{BCH}} \times R_{c_{LDPC}} \times \left[N_{OFDM_SYMB} - N_{P2} \right] \times N_{SCF_USED} + N_{P2} \times C_{P2} \right]$$
(31)
×(1 - OHF_{PP}) × T_U_FRACT] × [(N_{OFDM_SYMB} - N_{P2}) \times N_{SCF_USED} + N_{P2} \times C_{P2}]

На основе полученных выражений (1) ÷ (31) определим глубину перемежения OFDMсимволов и скорость передачи пользовательской информации для следующих условий [5]:

- полоса частот 8 МГц;
- $FFT_{SIZE} = 32KE;$
- $N_{b LDPC} = 64800;$
- $M_{mod} = 256;$
- $N_{P2} = 1;$
- $C_{P2} = 22432;$
- $C_{data} = 27404;$
- $T_U = 3.584 \text{ Mc};$
- GIF = 1/128;
- $N_{SCF} = 27840;$
- $R_{C LDPC} = 3/5;$

- $T_{P1} = 0.224 \text{ mc};$
- $T_U = 3.584$ Mc;
- $(1 OHF_{BB \ FRAME}) = 0.9985;$
- $R_{C BCH} = 0.997;$
- $(1 OHF_{PP7}) = 0.9896;$
- $T_{U FRACT} = 0.9922.$

Максимальное количество OFDM-символов в одном блоке временного перемежения в соответствии с (27) равно:

$$N_{OFDM} _SYMB} _MAX _IN _TI = \frac{N_C _MAX _TI}{N_{SCF} _in _OFDM} _SYMB} = \frac{557056}{27840} = 20.009 \Longrightarrow 20.$$
(32)

Максимальное количество блоков FEC в одном блоке временного перемежения в соответствии с (6) равно:

$$N_{BLOCK_FEC_MAX_IN_TI} = \frac{N_{C_MAX_TI}}{C_{BLOCK_FEC}} = \frac{557056}{8100} = 68.772 \Longrightarrow 68.$$
(33)

Количество строк в одном блоке временного перемежения в соответствии с (3) равно:

$$N_{ROW_IN_TI} = \frac{C_{BLOCK_FEC}}{5} = \frac{8100}{5} = 1620.$$
(34)

В [1, 5] для рассматриваемых условий выбрана длина кадра T2, содержащая $L_F = 60$ OFDM-символов. Для этого случая в соответствии с выражениями (26) и (27) получим необходимое число блоков временного перемежения в кадре T2:

$$N_{TI_in_T2F} = \frac{L_F}{N_{OFDM_SYMB_MAX_IN_TI}} = \frac{60}{20} = 3.$$
(35)

При длине кадра T2, содержащего $L_F = 55$ OFDM-символов, получим необходимое число блоков временного перемежения:

$$N_{TI_in_T2F} = \frac{L_F}{N_{OFDM_SYMB_MAX_IN_TI}} = \frac{55}{20} = 2.75 = 3.$$
 (36)

При длине кадра T2, содержащего $L_F = 65$ OFDM-символов, получим необходимое число блоков временного перемежения:

$$N_{TI_in_T2F} = \frac{L_F}{N_{OFDM_SYMB_MAX_IN_TI}} = \frac{65}{20} = 3.25 = 4.$$
 (37)

При этом в каждом блоке временного перемежения ТІ при заданном числе символов будет содержаться следующее количество OFDM-символов:

$$N_{OFDM} _ SYMB _ IN _ TI = \frac{L_F}{N_{TI} _ in _ T2F}.$$
(38)

В соответствии с (38) количество OFDM-символов в одном блоке временного перемежения TI при заданном числе символов L_F в кадре T2:

$$L_{F} = 55, N_{OFDM} _ S_IN_TI = 18; L_{F} = 60, N_{OFDM} _ S_IN_TI = 20; L_{F} = 65, N_{OFDM} _ S_IN_TI = 16$$
(39)

Полученное в (39) количество OFDM-символов в одном блоке временного перемежения TI *определяет глубину перемежения* этих символов.

Теперь определим количество фиктивных ячеек, которое будет создано во всех блоках $N_{TL in T2F}$ временного перемежения в кадре T2.

Сначала определим с учётом выражений (32) и (33) количество блоков FEC в одном блоке временного перемежения TI:

$$N_{BLOCK_FEC_IN_TI} = \frac{N_{BLOCK_FEC_MAX_IN_TI} \times N_{OFDM_SYMB_IN_TI}}{N_{OFDM_SYMB_MAX_IN_TI}}.$$
 (40)

В соответствии с (40) количество блоков FEC в одном блоке временного перемежения TI при заданном числе символов L_F в кадре T2 будет равно:

$$L_F = 55, N_{FEC_IN_TI} = 61; L_F = 60, N_{FEC_IN_TI} = 68; L_F = 65, N_{FEC_IN_TI} = 54.$$
(41)

Затем определим число столбцов в одном блоке временного перемежения TI (4) с учётом (40) и (41):

$$L_F = 55, N_{COL_IN_TI} = 305; L_F = 60, N_{COL_IN_TI} = 340; L_F = 65, N_{COL_IN_TI} = 275.$$
 (42)

Заметим, что число строк (ячеек) в блоке временного перемежения постоянно и в соответствии с (34) равно 1620.

Определим количество фиктивных ячеек при $L_F = 60$ OFDM-символов в кадре T2. При этом в кадре T2 содержится три блока временного перемежения TI (35), в каждом из трёх блоков содержится по 20 OFDM-символов (39). В одном TI-блоке 1620 строк (ячеек) (34) и 340 столбцов (42). Таким образом, в одном блоке временного перемежения всего 550800 ячеек и, следовательно, 68 блоков FEC. Всего в кадре перемежения при трёх TI-блоках будет содержаться 204 блока FEC (42). При этом общее количество блоков FEC в T2-кадре в соответствии с выражениями (16) и (18) равно:

$$N_{BLOCKFECT2F} = \frac{N_{P2} \cdot C_{P2} + L_{data} \cdot C_{data}}{C_{BLOCK} \quad FEC} = \frac{(1 \cdot 22432 + 59 \cdot 27404)}{8100} = 202.3787.$$
(43)

Следовательно, количество фиктивных ячеек в трёх блоках временного перемежения с учётом (28) и (43) будет определяться дробной частью общего количество блоков FEC в Т2кадре

$$C_{dummy} = \left(N_{BLOCK_FEC_IN_T2F} - N_{BLOCK_FEC_IN_T2F}\right) \times C_{BLOCK_FEC}.$$
 (44)

Тогда с учётом (43) и (44) количество фиктивных ячеек в трёх блоках временного перемежения будет равно:

$$C_{dummv} = 0.3787 \times 8100 = 3067.$$
⁽⁴⁵⁾

В соответствии с выражениями (29) и (43) – (45) относительное снижение числа информационных ячеек или скорости передачи из-за наличия фиктивных ячеек равно:

$$\Delta C_{T2F} = \frac{N_{P2} \cdot C_{P2} + L_{data} \cdot C_{data} - C_{dummy}}{N_{P2} \cdot C_{P2} + L_{data} \cdot C_{data}} = \frac{1639268 - 3067}{1639268} = 0.9982.$$
(46)

Таким образом, снижение скорости составит 0.18 %. Из формулы (31) следует, что без учёта фиктивных ячеек скорость пользовательских данных R_b wiTHOUT dummy = 36.288 Mбит/c, а с учётом использования фиктивных ячеек – R_b wiTH dummy = 36.222 Mбит/c.

Результаты расчётов влияния фиктивных ячеек на изменение скорости передачи для трёх значений $L_F = 55,\,60$ и 65 символов приведены в табл. 6.

$L_{\rm F}$	N _{TI}	N _{SYMB_IN_TI}	NFEC_BL_IN_T2F	N _{FEC_BL_IN_NTI}	C _{dummy}	ΔR_{b_UP}
54	3	18	182.0795	183	64	0.000043
60	3	20	202.3787	204	3067	0.0018
64	4	16	215.9116	216	7383	0.0042

Таблица 6. Результаты расчётов влияния фиктивных ячеек на изменение скорости передачи

Из табл. 6 следует, что использование фиктивных ячеек практически не изменяет скорость передачи из-за их малого количества.

На рис. 8 приведены результаты расчётов глубины перемежения (график 1) и скорости передачи (график 2) от длины кадра Т2. Из этого рисунка видно, что выбор длины кадра с L_F = 60 OFDM-символов даёт практически максимальную скорость передачи битов, а также максимизирует глубину перемежения, т.е. повышает помехоустойчивость приёма.



Рис. 8. Зависимость глубины перемежения (1) и скорости передачи (2) от длины кадра T2 (полоса частот 8 МГц; FFT_{SIZE} = 32KE; M_{mod} = 256; T_U = 3.584 мс; N_{SCF} = 27840; N_{P2} = 2; R_{C_LDPC} = 3/5)

На основе полученных выражений (1) ÷ (31) определим глубину перемежения OFDMсимволов и скорость передачи пользовательской информации для других сочетаний параметров стандарта DVB-T2 [5]:

- полоса частот 8 МГц;
- $FFT_{SIZE} = 8KN;$
- $N_{b_LDPC} = 64800;$

- $M_{mod} = 64;$
- $N_{P2} = 2;$
- $C_{P2} = 4472;$
- $T_U = 0.896$ Mc;
- GIF = 1/4;
- $N_{SCF USED} = 6816;$
- $R_{C_{LDPC}} = 3/4;$
- $T_{P1} = 0.224 \text{ mc};$
- $(1 OHF_{BB_{FRAME}}) = 0.9983;$
- $R_{C BCH} = 0.996;$
- $(1 OHF_{PP7}) = 0.9896;$
- $T_{U FRACT} = 0.8$.

Результаты проведённых расчётов для второго сочетания параметров стандарта DVB-T2 приведены в табл. 7 и на рис. 9.

Таблица 7. Зависимость глубины перемежения и скорости передачи от длины кадра L_F

L_F	N _{TI}	N _{SYMB} in ti	R _{b_UP} , Мбит/с	ΔR_{b} up
76	4	19	20.099	0.000546
80	4	20	20.110	0.0000
85	5	17	20.127	0.000844



Рис. 9. Зависимость глубины перемежения (1) и скорости передачи (2) от длины кадра T2 (полоса частот 8 МГц; FFT_{SIZE} = 8KN; M_{mod} = 64; N_{P2} = 2; T_U = 0.896 мс; N_{SCF} = 6816; $R_{C\ LDPC}$ = 3/4)

Из табл. 7 и рис. 9 видно, что для данного сочетания параметров выбор длины кадра с $L_F = 80$ OFDM-символов даёт практически максимальную скорость передачи битов и при этом максимизирует глубину перемежения, т.е. повышает помехоустойчивость приёма.

Разработанная методика определения оптимальной длины кадра в стандарте наземного цифрового телевизионного вещания DVB-T2 может быть использована для выбора оптимальных параметров систем связи, в которых используется технология ортогонального частотного мультиплексирования, например, WiMAX, LTE, Wi-Fi.

Литература

- 1. ETSI EN 302 755 V1.3.1 (2011-11) Digital Video Broadcasting (DVB); Frame structure channel coding and modulation for a second generation digital terrestrial television broadcasting system (DVB-T2).
- 2. ETSI TS 102 831 V1.2.1 (2012-08) Digital Video Broadcasting (DVB); Implementation guidelines for a second generation digital terrestrial television broadcasting system (DVB-T2).
- 3. *Носов В. И.* Обработка сигналов при ортогональном частотном мультиплексировании: учебное пособие. Новосибирск: СибГУТИ, 2012. 352 с.
- 4. Шахнович И. В. DVB-T2 новый стандарт цифрового телевизионного вещания // Связь и телекоммуникации. 2009. № 6. С. 30–35.
- 5. *Носов В. И., Зиновьев М. В., Ладан А. С.* Особенности определения параметров цифрового телевизионного вещания // Вестник СибГУТИ. 2021. № 2. С. 71–83.

Статья поступила в редакцию 11.04.2022; переработанный вариант – 16.05.2022.

Носов Владимир Иванович

д.т.н., профессор кафедры ЦТРВ и СРС СибГУТИ (630102, Новосибирск, ул. Кирова, 86), тел. (383) 2-698-254, e-mail: nvi@sibsutis.ru.

Ладан Александра Сергеевна

аспирант кафедры ЦТРВ и СРС СибГУТИ, e-mail: lad al@mail.ru.

Зиновьев Максим Вячеславович

аспирант кафедры ЦТРВ и СРС СибГУТИ, e-mail: neo136@list.ru.

Determination of the optimal frame length of digital television broadcasting

Vladimir I. Nosov

Doctor of technical sciences, Professor, Siberian State University of Telecommunications and Information Science (SibSUTIS, Novosibirsk, Russia), nvi@sibguti.ru.

Aleksandra S. Ladan

Postgraduate student, Siberian State University of Telecommunications and Information Science (SibSUTIS, Novosibirsk, Russia), lad_al@mail.ru.

Maksim V. Zinoviev

Postgraduate student, Siberian State University of Telecommunications and Information Science (SibSUTIS, Novosibirsk, Russia), neo136@list.ru.

The paper proposes a simple method for determining the optimal frame length in the DVB-T2 standard based on simple and understandable formulas based on the sequence of processing user data in the DVB-T2 standard. There should always be an integer number of FEC blocks in a T2 frame, the remaining cells that do not make up an entire FEC block will be dummy cells as a rule and represent a speed loss. As a result, the total bit rate and the depth of interleaving in the smallest details depend on the exact number of OFDM characters, and consequently, the cells in the T2 frame.

Keywords: DVB-T2 standard, T2-frame, OFDM symbol, FEC block, dummy cells, interleaving depth, optimal frame length.

References

- 1. ETSI EN 302 755 V1.3.1 (2011-11) Digital Video Broadcasting (DVB); Frame structure channel coding and modulation for a second generation digital terrestrial television broadcasting system (DVB-T2).
- 2. ETSI TS 102 831 V1.2.1 (2012-08) Digital Video Broadcasting (DVB); Implementation guidelines for a second generation digital terrestrial television broadcasting system (DVB-T2).
- 3. Nosov V.I. *Obrabotka signalov pri ortogonal'nom chastotnom mul'tipleksirovanii* [Signal processing with orthogonal frequency multiplexing]. SibGUTI, Novosibirsk, 2012. 352 p.
- 4. SHahnovich I.V. DVB-T2 novyj standart cifrovogo televizionnogo veshchaniya [DVB-T2 a new standard for digital television broadcasting]. *Svyaz' i telekommunikacii*, 2009, no. 6, pp. 30-35.
- Nosov V.I., Zinov'ev M.V., Ladan A.S. Osobennosti opredeleniya parametrov cifrovogo televizionnogo veshchaniya [Features of determining the parameters of digital television broadcasting]. *Vestnik SibGUTI*, 2021, no. 2, pp. 71-83.

Сбор данных о работе оборудования в сети мобильного оператора связи

С. Б. Жанаева

Операторы мобильной сети передачи данных с ростом количества оборудования сталкиваются с увеличивающейся сложностью эксплуатации и растущими затратами на обслуживание. При увеличении количества базовых станций растет и количество сбоев. Современные технологические решения, основанные на алгоритмах нейронных сетей, способны заблаговременно с определенной вероятностью предсказать возникновение сбоев в работе оборудования. Для обучения модели нейронной сети требуются данные о работе и сбоях на оборудовании мобильной сети передачи данных. В данной статье рассказывается о выполненном сборе данных в сети мобильного оператора 4G+, об особенностях и ограничениях, которые в дальнейшем могут повлиять на обучение модели.

Ключевые слова: сбор данных, мобильные сети, сбои в работе оборудования, центр управления сетью.

1. Введение

С развитием технологии виртуализации сетей и внедрением 5G/IoT мобильные сети предлагают более разнообразные услуги и становятся более сложными. Большинство телекоммуникационных мобильных операторов контролируют сотни тысяч микроволновых каналов различной конфигурации. Как следствие, центры управления сетью (Network Operational Center, NOC) часто бывают перегружены поступающими данными о состоянии базовых станций и мгновенными сигналами аварий или сбоев в работе оборудования.

Несмотря на то, что центры управления сетью NOC могут выполнять множество задач, увеличивающийся размер мобильных сетей и большое количество разнообразных аварийных сигналов затрудняют обслуживание сети и делают его более ресурсозатратным [1].

К тому же нередко случается, что на сети происходят тысячи аварийных сигналов в минуту, и перед сотрудниками центра управления сетью появляется несколько задач, например, им необходимо вручную проанализировать и определить приоритетные базовые станции, на которые следует в первую очередь отправить инженерную бригаду, и распознать/предположить вероятную причину проблемы на станциях.

Вследствие этого центрам управления сетью зачастую приходится отправлять сервисные группы на места для замены оборудования, не зная точной причины проблемы. Это часто приводит к ненужным дорогостоящим посещениям объекта и замене оборудования, возможно, без обнаружения неисправности.

Специалисты, работающие в центрах управления сетью NOC, собирая и анализируя большой объем данных, охватывающий определённый промежуток времени, с высокой долей вероятности могут быстро и легко найти первопричину неисправности в микроволновой линии связи. Но постоянная ручная работа по перебору большого объема данных в сетях с десятками тысяч микроволновых линий связи с целью найти и определить небольшую выборку проблемных станций является сложной задачей, требующей постоянной большой внимательности и определенной нагрузки на зрение.

Задачи по работе с большими объемами данных как раз являются той областью, где алгоритмы глубокого обучения могут принести немалую пользу и выполнить утомительную работу [2].

Можно обучить алгоритмы решать как сложные, так и более простые задачи, выполняемые экспертами, а также возложить на них задачу непрерывно просматривать огромные объемы данных в поисках первопричины сбоев. Это возможность использования алгоритмов машинного обучения с целью помочь операторам связи управлять сетевой инфраструктурой более эффективно по времени и с меньшими эксплуатационными расходами.

2. Обзор литературы

Со времени увеличения мощностей вычислительных машин и последовавшего затем возрождения и успехов алгоритмов машинного и глубокого обучения многие разработчики заинтересовались темой применения искусственного интеллекта в телекоммуникационной сфере, так как именно эта область деятельности человека сопровождается производством огромного количества структурированных данных. В том числе многих исследователей заинтересовал вопрос прогнозирования сбоев в сетях мобильного оператора, и ими был выполнен сбор необходимых для исследования данных.

Например, авторы в [3-5] собирали данные об авариях на базовых станциях, получаемых ежедневно, объединяли их с диспетчерской информацией о неисправностях на базовых станциях с целью установления связи между временем возникновения неисправности и предупреждающей информацией об аварии за *К* дней до возникновения неисправности. В качестве источника входных данных они использовали тревожную информацию с базовых станций (Alarms), генерируемую модулем анализа обработки тревожных сигналов, а в качестве источника данных маркировки использовали диспетчерскую информацию о неисправностях базовых станций (Tickets). Авторы в [3-5] описали интеллектуальную систему управления мобильной сетью, созданную с помощью методов контролируемого машинного обучения в целях создания системы интеллектуального обслуживания.

Авторы в [6, 7] предлагают реализовать систему прогнозирования сбоев и мониторинга сервиса, способную прогнозировать и оценивать качество услуг, близкое к тому, что на самом деле испытывают клиенты, оценивая его по сетевым данным состояния базовых станций и данным по авариям. Для достижения цели они предлагают собирать различные данные из сети, включая информацию об аварийных сигналах оборудования, и объединять их в большие кластеры данных для многовекторного анализа с помощью алгоритмов глубокого обучения. При реализации сервисного мониторинга авторы предлагают преобразовать процесс технического обслуживания из технического обслуживания объекта, сосредоточенного на аварийных сигналах оборудования, в техническое обслуживание сети, основанное на опыте клиентов.

Авторы в работе [8] для улучшения обслуживания и прогнозирования сбоев предлагают использовать собранную информацию об авариях на базовых станциях и данные о сроках эксплуатации оборудования, увеличивая расчетную вероятность отказа оборудования с возрастанием срока эксплуатации.

Большинство авторов [3–10] сходятся во мнении о необходимости использования нескольких типов источников данных. Так, например, авторы [10, 11] советуют, помимо данных мониторинга состояния, также собирать и использовать в прогнозировании сбоев влияние переменных внешней среды, таких как данные о рабочем состоянии, а также влиянии незначительных действий по техническому обслуживанию.

3. Описание собранных данных

Для того чтобы обучить алгоритмы машинного обучения и помочь операторам связи в управлении сетевой инфраструктурой, необходимо было собрать требуемые данные о работе сети. На данном этапе проведения исследовательской работы был выполнен сбор данных по качеству работы сетевого оборудования, базовых станций, по количеству и времени вынужденных простоев. Сбор данных проводился в течение нескольких месяцев, в разные дни недели и в разное время суток. Дни недели и время для сбора статистических данных определялись прогнозируемым объемом трафика и зависели от степени загруженности сети. Все данные сохранялись специальным программным обеспечением в формате *csv*-файлов.

3.1. Данные по авариям на оборудовании

В первую очередь необходимо было собрать данные, выдаваемые системой о различных аварийных событиях в сети, с учетом определенного специальным программным обеспечением по наблюдению за станциями распределения аварий по уровню критичности. При эксплуатации беспроводной сети такого типа данные об аварийных ситуациях собираются на постоянной основе и сохраняются в дневных, недельных, месячных отчетах (Alarms Daily Report, Alarms Weekly Report, Alarms Monthly Report) и в дальнейшем анализируются несколькими отделами компании оператора связи. Всего были собраны данные по аварийным ситуациям на 2700 станциях, данные по каждой станции и сектору были сохранены под определённым идентификатором. Сбор аварийных данных проводился в течение 10 месяцев, был собран большой объем данных на основе ежедневного отчета об аварийных сигналах Alarms Daily Report, который содержит такие параметры, как дата, время, станция, сектор, тип аварии, наименование аварии. Данные сохранены в формате *csv*-файлов и содержат около 3 млн записей (в дальнейшем часть записей была удалена из базы данных как дублированная).

3.2. КРІ-показатели оборудования

Были собраны и сохранены данные о текущем состоянии базовых станций для того, чтобы в дальнейшем проанализировать их и определить наиболее вероятную первопричину сбоев на сети. Эти данные характеризуют производительность сетевого оборудования и отображают основные показатели KPI (Key Performance Indicator). Было собрано два типа подобных отчетов:

1. Показатели Average KPI (среднее значение показателей производительности) базовых станций: собрана информация усредненных за день технических показателей качества сервисов для каждой соты, на которой предоставляли сервис абонентам.

2. Показатели Channel KPI (канальное значение показателей производительности) базовых станций: собрана информация об изменениях (в час наибольшей нагрузки) технических показателей качества сервисов для каждой соты, на которой предоставляли сервис абонентам.

Каждый из двух вышеперечисленных отчетов по состоянию 2700 базовых станций содержит 42 показателя, которые в дальнейшем планируется использовать в качестве характеристик данных и атрибутов при обучении алгоритмов машинного обучения. Сбор данных проводился в течение 10 месяцев, и каждый из отчетов Average KPI и Channel KPI содержит более 2 миллионов записей.

3.3. Причины сбоев в работе оборудования

Системы, отображающие состояние сети и базовых станций, по умолчанию выдают предположительную причину аварийного состояния, но бывает так, что предложенная системой причина не соответствует действительности. Например, выдаваемая системой авария Rectifier Failure (отказ выпрямителя) зачастую на практике означает выход из строя кондиционера и перегрев на станции. Поэтому одновременно со сбором данных по аварийным ситуациям на сети также был проведен сбор данных по выяснившимся первопричинам аварий. Первопричины аварий и сбоев на сети были выявлены и проверены выехавшими на место сетевыми инженерами и инженерами связи. Инженерный персонал ежедневно созванивался и высылал в центр управления сетью отчеты по устраненным авариям и сбоям на сети с выявленной и подтверждённой причиной и перечнем предпринятых ими действий для устранения аварии. Таким образом, была сформирована база данных по первопричинам аварийных ситуаций на сети.

После завершения сбора данных из общего объема трех вышеперечисленных типов данных была сделана выборка статистически значимого объема данных по анализируемым станциям.

3.4. Сложности в сборе данных

По признаку «видимости» сбои на оборудовании базовых станций мобильного оператора условно можно разделить на два основных типа: видимые явные и невидимые скрытые. Видимые или явные сбои в работе оборудования отображаются приходящими с системы управления и мониторинга сигналами тревоги, аварии на оборудовании и высвечиваются в соответствующих окнах GUI. Увидев их, сотрудники центра мониторинга могут немедленно связаться с инженерной бригадой и отправить ее на место для решения проблемы в работе оборудования.

Иногда кроме видимых аварий, явно отображаемых в системе мониторинга, возникают так называемые невидимые скрытые аварии в сети. Например, иногда на некоторых базовых станциях можно заметить полное пропадание трафика и возникновение молчания, на станцию как будто не поступают запросы на соединение от абонентов, кажется, что на данной станции или в секторе абоненты отсутствуют. Такая ситуация иногда возникает и бывает вызвана «зависанием» программного обеспечения станции [3]. Сотрудники службы мониторинга могут заметить отсутствие трафика и зависание программного обеспечения, для решения проблемы они обычно выполняют операцию Reset станции, и через минуту-две трафик на станции восстанавливается. При этом в системе мониторинга возникает авария отключения станции, падения радиоканала. Как следствие, данная авария сохраняется в базе данных в виде аварии отключения базовой станции и в дальнейшем при обучении нейронной модели прогнозированию сбоев может быть неверно интерпретирована и ошибочно воспринята как авария отключения питания или помехи в радиоканале. Подобную аварию можно распознать, одновременно собирая данные по показателям производительности базовых станций, и выявить продолжительное отсутствие трафика перед отключением станции.

4. Результаты сбора данных

4.1. Данные по авариям на оборудовании

Всего было собрано три комплекта данных: база данных по авариям на оборудовании, база данных по состоянию базовых станций с показателями производительности KPI и база данных по первопричинам сбоев на станциях. Все данные были сохранены в формате *csv*-

файлов с разделительными запятыми. Данные были проанализированы с помощью библиотек Python: Numpy и Pandas, были определены количество собранных данных, размерность.

```
import numpy as np
import pandas as pd
url = '.../alarm_daily/data'
data_alrm = pd.read_csv(url, header=None)
data_alrm.shape ()
(3847653, 5)
```

Так, база данных по авариям состоит из 3847653 строк, каждая строка отображает аварийное состояние оборудования мобильной сети. Столбцы содержат следующие данные: уровень критичности аварии, наименование и ID базовой станции, время фиксирования аварии, предположительная причина аварии, наименование трансмиссии. В табл. 1 показана выборка и пример содержащейся информации в базе данных по авариям.

N⁰	Уровень	Наименование	Время	Причина	Трансмиссия
1	Minor	37323_AKT-Beyneu	2021-02-15 11:32:28	Abis control link broken	Aktau-BSC-2
2	Minor	51247_KZL-Merke	2021-02-15 11:32:28	Abis control link broken	Kzl-BSC-1
3	Major	71422_ALM-Issyk	2021-02-15 11:32:28	Rectifier failure	Alm-BSC-2
4	Critical	37375_AKT-Karasu	2021-02-15 11:32:29	POWER OFF	Aktau-BSC-2
5	Minor	71458_ALM-Issyk	2021-02-15 11:32:29	Abis control link broken	Alm-BSC-1

Таблица 1. Выборка из базы данных по аварийным ситуациям в сети

База данных по авариям обладает большой избыточностью, так как зачастую на одно аварийное состояние оборудования система мониторинга и записи аварий выдает несколько тревожных сообщений с разницей в секунды, в основном такое отображение аварийных состояний характерно при помехах в радиоканале. На рис. 1 показано избыточное отображение одной и той же аварии в системе мониторинга с разницей в несколько секунд.

Major	54836_SMP-Ayaguz	2021-03-18 16:21:42	Rectifier failure	Semey-BSC-1
Critical	54836_SMP-Ayaguz	2021-03-18 16:21:43	POWER OFF	Semey-BSC-1
Minor	72423_ALM-Issyk-3	2021-03-18 16:21:43	Abis control link broken	Alm-BSC-2
Minor	72423_ALM-Issyk-3	2021-03-18 16:21:45	Abis control link broken	Alm-BSC-2
Minor	72423_ALM-Issyk-3	2021-03-18 16:21:46	Abis control link broken	Alm-BSC-2
Minor	72423_ALM-Issyk-3	2021-03-18 16:22:10	Abis control link broken	Alm-BSC-2
Minor	72423_ALM-Issyk-3	2021-03-18 16:22:12	Abis control link broken	Alm-BSC-2
Minor	72423_ALM-Issyk-3	2021-03-18 16:22:14	Abis control link broken	Alm-BSC-2

Рис. 1. Дублирующее отображение аварий в системе мониторинга

В дальнейшем при подготовке данных из базы данных по авариям избыточность была удалена.

4.2. Данные по состоянию базовых станций

Был проведен сбор данных по состоянию базовых станций. Как было описано в пункте 3.2, состояние базовых станций характеризуется двумя наборами данных: показателями Average KPI и показателями Channel KPI.

Информация, содержащаяся в этих комплектах данных, необходима для анализа состояния базовых станций до момента появления аварии, чтобы в дальнейшем проанализировать и найти возможную корреляционную зависимость между определенными авариями и изменениями показателей состояния оборудования базовых станций с целью прогнозирования первопричин аварий.

В каждой из собранных баз данных Average KPI и Channel KPI содержится информация по 2700 базовым станциям и 42 показателям. С помощью библиотек Python Numpy и Pandas были определены количество собранных данных и размерность.

```
import numpy as np
import pandas as pd
url_a = '.../average_kpi/data'
url_c = '.../channel_kpi/data'
data_average_kpi = pd.read_csv(url_a, header=None)
data_channel_kpi = pd.read_csv(url_c, header=None)
data_average_kpi.shape ()
data_channel_kpi.shape ()
(896727, 42)
(896560, 42)
```

Всего было собрано около 900 тысяч данных по 42 показателями, по каждой базе данных Average KPI и Channel KPI. В целях прогнозирования сбоев и вероятностного определения первопричин аварий будут необходимы данные только по базовым станциям, на которых по тем или иным причинам возникали аварийные сообщения. Поэтому из всего объема собранных данных по показателям производительности базовых станций будут выбраны только данные с номерами базовых станций, совпадающими с номерами станций из базы данных по аварийным сообщениям. Таким образом, базы данных Average KPI и Channel KPI будут уменьшены.

4.3. Данные по первопричинам сбоев

В результате совместной работы сотрудников центра мониторинга сети и инженерного состава, обслуживающего оборудование базовых станций, была сформирована база данных по первопричинам аварийных ситуаций на сети. Эти данные необходимы для обучения модели прогнозирования и определения первопричин аварийных ситуаций в работе оборудования. Собранные данные были скомпонованы сотрудниками отдела эксплуатации базовых станций и сохранены в формате *csv*-файлов. С помощью библиотек Python Numpy и Pandas были определены количество собранных данных и размерность.

```
import numpy as np
import pandas as pd
url_pa = '.../alrm_reasons/data'
data_alrm_reasons = pd.read_csv(url_alrm_reasons, header=None)
data_alrm_reasons.shape ()
(68596, 4)
```

Было подсчитано, что база данных с первопричинами аварий состоит из 68596 строк. Столбцы базы данных содержат следующую информацию: наименование и ID базовой станции, время фиксирования аварии, выявленная причина аварии, трансмиссия.

5. Заключение

На данном этапе исследовательской работы был проведен сбор данных по работе оборудования базовых станций мобильной сети телекоммуникационного оператора 4G+. Сбор данных проводился с целью дальнейшего использования для разработки нейронной модели прогнозирования сбоев в работе оборудования мобильной сети. Разработанная модель призвана послужить в качестве решающего средства для предотвращения непредсказуемых потерь в работе сети, оптимизации расходов на обслуживание оборудования, способствуя тем самым повышению качества предоставляемых услуг.

Литература

- 1. Тихвинский В. О., Терентьев С. В., Коваль В. А. Сети мобильной связи 5G: технологии, архитектура и услуги. М.: Медиа Паблишер, 2019. 375 с.
- 2. Brink H., Richards J. W., Fetherolf M. Real world Machine Learning. US Manning Publications Co., 2017. 266 p.
- Yin-Hsin Liu, Yao-Chung Tu, Chang-Yu Hsu, Hsin-Chieh Chao. Predicting malfunction of mobile network base station using machine learning approach // Proc. 20th Asia-Pacific Network Operations and Management Symposium (APNOMS), 18–20 September, Matsue, Japan. 2019. P. 1–4. DOI: 10.23919/APNOMS.2019.8892894.
- Kumar Y., Farooq H., Imran A. Fault Prediction and Reliability Analysis in a Real Cellular Network // Proc. 13th International Wireless Communications and Mobile Computing Conference (IWCMC), Valencia, Spain, 2017. P. 1090–1095. DOI:10.1109/IWCMC.2017.7986437.
- Corazza A., Isgro F., Longobardo L., Prevete R. A machine learning approach for predictive maintenance for mobile phones service providers // Advances on P2P, Parallel, Grid, Cloud and Internet Computing / Lecture Notes on Data Engineering and Communications Technologies. / F. Xhafa (Ed.). Springer International Publishing, 2017. DOI: 10.1007/978-3-319-49109-7_69.
- 6. *Otani T., Toube H., Kimura T., Furutani M.* Application of AI to mobile network operation // ITU Journal: ICT Discoveries. 2018. Special Issue. № 1. P. 43–48.
- Hu C., Youn B. D., Kim T. Semi-supervised learning with co-training for data-driven prognostics // Proc. IEEE Int. Conf. on Prognostics and Health Management: Enhancing Safety, Efficiency, Availability, and Effectiveness of Systems Through PHM Technology and Application, Denver, CO, USA. 18–21 June 2012. P. 1297–1306. DOI: 10.1109/ICPHM.2012.6299526.
- 8. *Yousefia N., Tsianikasa S., Zhoua J. Coit D. W.* Inspection plan prediction for multi-repairable component systems using neural network // Proc. IISE Annual Conference. / L. Cromarty, R. Shirwaiker, P. Wang (Eds.). Institute of Industrial and Systems Engineers, IISE, Virtual, Online, United States. 2020.
- 9. Schmidt B., Wang L. Cloud-enhanced predictive maintenance // Int J Adv Manuf Technol. 2018. № 99. P. 5–13. DOI 10.1007/s00170-016-8983-8.
- 10. *Wei Wu, Feng Zhang, Min Liu, Weiming Shen.* A Multi-agent Based Failure Prediction Method Using Neural Network Algorithm // Proc. IEEE International Conference on Systems, Man, and Cybernetics. San Diego, CA, 2014. P. 2268–2272. DOI: 10.1109/SMC.2014.6974263.
- Karballaeezadeh N., Zaremotekhases F., Shamshirband S., Mosavi A., Nabipour N., Csiba P., Várkonyi-Kóczy A. Intelligent Road Inspection with Advanced Machine Learning; Hybrid Prediction Models for Smart Mobility and Transportation Maintenance Systems // Energies. 2020. № 13. P. 1718. DOI: 10.3390/en13071718.

Статья поступила в редакцию 03.05.2022; переработанный вариант – 17.05.2022.

Жанаева Сауле Бактыкереевна

аспирант, кафедра прикладной математики и кибернетики СибГУТИ (630102, Новосибирск, ул. Кирова, 86), e-mail: szhanayeva@gmail.com.

Data collection concerning equipment operation in the network of a mobile operator

Saule B. Zhanayeva

Postgraduate student, Siberian State University of Telecommunications and Information Science (SibSUTIS, Novosibirsk, Russia), szhanayeva@gmail.com.

Mobile data network operators face increasing operational complexity and rising maintenance costs as the number of equipment increases. As the number of base stations increases, so does the number of failures. Modern technological solutions based on neural network algorithms are able to predict in advance with a certain probability the occurrence of equipment failures. Data of operation and failures on the mobile network equipment is required to train a neural network model. The article considers the performed data collection on the 4G mobile operator network, features and limitations that may further affect the model training.

Keywords: data collection, mobile networks, equipment failures, network operational center.

References

- 1. Tihvinskiy V. O., Terentev S. V., Koval V. A. *Seti mobilnoy cvyazi 5G: tehnologii, arhitektura i uslugi* [Mobile Networks 5G: technologies, architecture and services]. Moscow, Media Publishers, 2021, 375 p.
- 2. Brink H., Richards J. W., Fetherolf M. *Real world Machine Learning*. US, Manning Publications Co., 2017, 266 p.
- 3. Yin-Hsin Liu, Yao-Chung Tu, Chang-Yu Hsu, Hsin-Chieh Chao. Predicting malfunction of mobile network base station using machine learning approach. *Proceedings of the 20th Asia-Pacific Network Operations and Management Symposium (APNOMS)*, 18-20 September, Matsue, Japan, 2019, pp. 1-4. DOI: 10.23919/APNOMS.2019.8892894.
- 4. Kumar Y., Farooq H., Imran A. Fault Prediction and Reliability Analysis in a Real Cellular Network. *Proceedings of 13th International Wireless Communications and Mobile Computing Conference* (*IWCMC*), 26-30 June, pp. 1090-1095, Valencia, Spain, 2017. DOI: 10.1109/IWCMC.2017.7986437.
- Corazza A., Isgro F., Longobardo L., Prevete R. A machine learning approach for predictive maintenance for mobile phones service providers. *Advances on P2P, Parallel, Grid, Cloud and Internet Computing. Lecture Notes on Data Engineering and Communications Technologies,* Springer International Publishing, 2017. DOI: 10.1007/978-3-319-49109-7_69.
- 6. Otani T., Toube H., Kimura T., Furutani M. Application of AI to mobile network operation. *ITU Journal: ICT Discoveries*, 2018, Special Issue, no. 1, pp. 43-48.
- Hu C., Youn B. D., Kim T. Semi-supervised learning with co-training for data-driven prognostics. Proceedings of IEEE Int. Conf. on Prognostics and Health Management: Enhancing Safety, Efficiency, Availability, and Effectiveness of Systems Through PHM Technology and Application, 18-21 June, Denver, CO, USA, 2012, pp. 1297-1306. DOI: 10.1109/ICPHM.2012.6299526.
- 8. Yousefia N., Tsianikasa S., Zhoua J., Coit D. W. Inspection plan prediction for multi-repairable component systems using neural network. *Proceedings of the 2020 IISE Annual Conference*, Institute of Industrial and Systems Engineers, IISE, United States, 1-3 Nov, 2020.
- 9. Schmidt B., Wang L. Cloud-enhanced predictive maintenance. *Int J Adv Manuf Technol*, 2018, no. 99, pp. 5–13. DOI 10.1007/s00170-016-8983-8.
- 10. Wei Wu, Feng Zhang, Min Liu, Weiming Shen. A Multi-agent Based Failure Prediction Method Using Neural Network Algorithm. *Proceedings of 2014 IEEE International Conference on Systems, Man, and Cybernetics,* San Diego, CA., 2014, pp. 2268-2272, DOI: 10.1109/SMC.2014.6974263.
- Karballaeezadeh N., Zaremotekhases F., Shamshirband S., Mosavi A., Nabipour N., Csiba P., Várkonyi-Kóczy A. Intelligent Road Inspection with Advanced Machine Learning; Hybrid Prediction Models for Smart Mobility and Transportation Maintenance Systems. *Energies*, 2020, no. 13, pp. 1718. DOI: 10.3390/en13071718.

Оценка помехоустойчивости DWDM-сигналов в нелинейном режиме функционирования волоконно-оптического тракта

А. Н. Сычук, В. А. Варданян

Исследуется совместное влияние явлений фазовой самомодуляции, фазовой кроссмодуляции, четырёхволнового смешения и шума усиленной спонтанной эмиссии на помехоустойчивость передаваемых импульсных сигналов в многоканальной волоконнооптической системе передачи со спектральным уплотнением в зависимости от параметров системы передачи: уровня суммарной мощности оптического излучения в оптоволокне, скорости передачи и количества спектральных каналов. Критерием оценки помехоустойчивости импульсных сигналов принят *Q*-фактор. Показана зависимость величины суммарного *Q*-фактора от уровня суммарной мощности излучения в оптоволокне для разных значений скорости передачи и количества спектральных каналов.

Ключевые слова: помехоустойчивость импульсных сигналов, спектральное разделение каналов, фазовая кросс-модуляция, четырёхволновое смешение, шум усиленной спонтанной эмиссии.

1. Введение

Технология частотного разделения каналов в оптическом диапазоне WDM (Wavelengths Division Multiplexing) повсеместно внедрена в современные волоконно-оптические системы передачи (ВОСП) и обеспечивает высокую эффективность использования пропускной способности оптического волокна. Принцип работы технологии WDM заключается в одновременной передаче информационных сигналов с помощью большого количества спектрально разделённых каналов. Частотный интервал между спектральными каналами регламентирован стандартом ITU-T G. 694.1 [1]. Увеличение количества спектральных каналов является одним из способов повышения пропускной способности ВОСП-WDM. Однако большое число спектральных каналов приводит к возрастанию суммарной мощности оптического излучения в оптоволокне, что, в свою очередь, приводит к возникновению различных нелинейных явлений типа фазовой самомодуляции (ФСМ) и фазовой кросс-модуляции (ФКМ), а также четырёхволнового смешения (ЧВС) [2, 3]. Явление ЧВС в волокне приводит к генерации комбинационных частот [4-6]. Некоторые из комбинационных частот попадают в диапазоны длин волн спектральных каналов, приводя к возникновению перекрёстных помех и снижению качества передаваемых сигналов [7]. Количество генерируемых комбинационных частот увеличивается с ростом количества спектральных каналов, тем самым усиливая влияние явления ЧВС на качество передаваемых сигналов [8]. В свою очередь, явления ФСМ и ФКМ влияют на фазы сигналов передаваемых спектральных каналов. В процессе фотодетектирования искажение фазы сигналов в спектральных каналах трансформируется в искажение амплитуды фотодетектируемых сигналов, что приводит к снижению помехоустойчивости принимаемых сигналов [2]. В современных ВОСП-WDM дальность передачи может быть увеличена путем включения оптических усилителей в оптический тракт. Наиболее распространённым типом усилителей, применяемых в ВОСП-WDM, является эрбиевый усилитель EDFA (Erbium Doped Fiber Amplifier). Однако усилитель EDFA является источником шума спонтанной эмиссии ASE (Amplified Spontaneous Emission). ASE-шум возникает в процессе когерентного фотодетектирования при смешивании детектируемого сигнала со спонтанным шумом усилителя [2, 6].

Целью данной работы является исследование совместного влияния явлений ФСМ, ФКМ, ЧВС и шума ASE на помехоустойчивость импульсных сигналов в многоканальной ВОСП-WDM с гомодинным приёмом. Критерием оценки помехоустойчивости импульсных сигналов принят *Q*-фактор.

2. Структурная схема системы передачи

На рис. 1 представлена структурная схема многоканальной ВОСП-WDM с гомодинным фотоприёмником.



Рис. 1. Структурная схема многоканальной ВОСП-WDM с когерентным детектированием

В передающей части лазерные диоды (ЛД) генерируют оптические несущие для спектральных каналов, диапазоны длин волн которых определены в соответствии с сеткой частот DWDM (Dense WDM). Оптические несущие генерируются в диапазоне С для стандартного одномодового оптического волокна (Standard Single Mode Fiber, SSMF). Для лазерных диодов предусмотрена возможность контроля уровня мощности в спектральном канале. В работе предполагается, что суммарная мощность, вводимая в оптоволокно, распределена равномерно между всеми спектральными каналами. Излучение от лазерных диодов поступает на один из входов оптических модуляторов (М). На другой вход оптических модуляторов подаются последовательности управляющих импульсов, формируемых генератором импульсов NRZ, RZ (ГИ NRZ, RZ). В генераторе импульсов NRZ, RZ предусмотрена возможность регулировки длительности управляющих импульсов. Отметим, что в процессе модуляции под логической «1» принято наличие излучения на выходе оптического модулятора, а под логическим «0» принято отсутствие оптического излучения на выходе оптического модулятора. Также в работе учтено наличие межсимвольной интерференции между оптическими импульсами в соседних тактовых интервалах. Модулированные оптические несущие мультиплексируются (MUX) и поступают на вход оптического усилителя (ОУ). Оптические усилители используются в схеме для компенсации потерь мощности, возникающих в передающей и линейной частях оптического тракта. В данной работе используются два EDFA-усилителя: бустерный усилитель - BOA (Booster Optical Amplifier) и предусилитель - OPA (Optical Pre-Amplifier). В процессе распространения мультиплексированного сигнала по оптоволокну на сигнал оказывают воздействие нелинейные явления: ФСМ, ФКМ и ЧВС. Также в модели учитывается влияние шума ASE. В приёмной части после предварительного усиления и демультиплексирования (DMUX) демультиплексированный сигнал поступает на вход фотоприёмного устройства (ФПУ), состоящего из направленного ответвителя (НО), местного гетеродина (МГ) и фотодетектора (ФД). Усиленный сигнал вводится в направленный ответвитель. В направленном ответвителе происходит смешение поля детектируемого сигнала и поля оптического излучения, генерируемого местным гетеродином. Результирующее поле из направленного ответвителя поступает на фотодетектор. На выходе фотодетектора формируется фототок, который регистрируется осциллографом (Осц), синхронизированным с генератором импульсов NRZ, RZ.

3. Теоретическое обоснование

В приёмной части ВОСП-WDM оптическое поле на входе фотодетектора представляет собой результат взаимодействия оптического поля гетеродина с оптическим полем детектируемого сигнала, искаженного в результате воздействия явлений ФСМ, ФКМ, ЧВС и оптического шума ASE. В результате огибающая фототока на выходе фотодетектора может быть описана следующим образом [2]:

$$I_{Pd}(t) = i_{NL}(t) + i_{S-ASE}(t) + i_{LO-ASE}(t),$$
(1)

где $i_{NL}(t)$ – флуктуация фототока из-за воздействия явлений ФСМ, ФКМ и ЧВС на детектируемый сигнал; $i_{S-ASE}(t)$ – флуктуация фототока из-за биения детектируемого сигнала со спонтанным излучением; $i_{LO-ASE}(t)$ – флуктуация фототока из-за биения излучения гетеродина со спонтанным излучением. Отметим, что в (1) параметр $i_{NL}(t)$ описывает суммарное воздействие явлений ФСМ, ФКМ и ЧВС вследствие их общего механизма возникновения, связанного со взаимодействием спектральных каналов из-за нелинейной восприимчивости оптоволокна третьего порядка [5]. Однако явление ЧВС и явления ФСМ и ФКМ по-разному оказывают влияние на помехоустойчивость детектируемого сигнала [3, 4]. Поэтому в дальнейшем оценка влияния явлений ФСМ, ФКМ и влияние явления ЧВС на помехоустойчивость детектируемого сигнала будет выполняться по отдельности.

Оценку влияния шума ASE на *Q*-фактор детектируемого сигнала удобно проводить с помощью оценки дисперсии шума ASE, которая определяется как результирующее значение дисперсии флуктуаций фототока из-за биения детектируемого сигнала со спонтанным излучением и дисперсии флуктуаций фототока из-за биения излучения гетеродина со спонтанным излучением. Влияние шума ASE на *Q*-фактор детектируемого сигнала можно определить следующим образом [3]:

$$Q_{ASE} = \frac{I_1 - I_0}{\sigma_{S - ASE} + \sigma_{LO - ASE}},\tag{2}$$

где I_1 – уровень фототока сигнала при передаче логической «1»; I_0 – уровень фототока сигнала при передаче логического «0»; σ_{S-ASE} – среднеквадратическое отклонение флуктуаций фототока из-за биения детектируемого сигнала со спонтанным излучением; σ_{LO-ASE} – среднеквадратическое отклонение флуктуации фототока из-за биения излучения гетеродина со спонтанным излучением. Дисперсию флуктуации фототока из-за биения детектируемого сигнала со спонтанным излучения гетеродина со спонтанным излучением. Дисперсию флуктуации фототока из-за биения детектируемого сигнала со спонтанным излучением можно определить следующим образом [2, 6]:

$$\sigma_{S-ASE}^2 = 4 \cdot v \cdot h \cdot \Delta f \cdot \mathbf{n}_{ASE} \cdot \mathbf{P}_{\ll 1}$$
(3)

Соответственно, дисперсия флуктуации фототока из-за биения излучения гетеродина со спонтанным излучением определяется как [3, 7]:

$$\sigma_{LO-ASE}^2 = 4 \cdot v \cdot h \cdot \Delta f \cdot \mathbf{n}_{ASE} \cdot \mathbf{P}_{LO} , \qquad (4)$$

где v – частота излучения детектируемого сигнала, равная частоте излучения гетеродина (когерентный гомодинный приём); h – постоянная Планка; Δf – ширина полосы частот детектированного сигнала; n_{ASE} – коэффициент спонтанного излучения; $P_{«1»}$ – уровень пиковой мощности оптического импульса детектируемого сигнала, соответствующего переданной логической «1»; P_{LO} – уровень мощности оптического излучения гетеродина.

Отметим, что дисперсия шума ASE зависит от уровней мощности сигнала гетеродина и детектируемого сигнала, однако при расчётах в [2] было установлено, что вклад дисперсии флуктуации фототока из-за биения детектируемого сигнала со спонтанным излучением на два порядка ниже по сравнению с дисперсией флуктуации фототока из-за биения излучения гетеродина со спонтанным излучением. По этой причине при дальнейших расчётах вклад дисперсии флуктуации фототока из-за биения излучением на на ересии флуктуации фототока из-за биения детектируемого сигнала со спонтанным излучения излучением. По этой причине при дальнейших расчётах вклад дисперсии флуктуации фототока из-за биения за биения детектируемого сигнала со спонтанным излучением не учитывается.

Оценку влияния явления ЧВС (Four Wave Mixing, FWM) на *Q*-фактор детектируемого сигнала удобно проводить в частотной области при помощи выражения [4–6]:

$$Q_{FWM} = \left[\frac{P_m}{P_{FWM}}\right]^{\frac{1}{2}},\tag{5}$$

где P_m – средний уровень мощности в *m*-м канале, m = 1, 2, ...N, N – количество спектральных каналов; P_{FWM} – суммарная мощность комбинационных частот, попадающих в *m*-й канал. Суммарную мощность комбинационных частот в *m*-м канале при длине оптического волокна *L* можно определить следующим образом:

$$P_{FWM} = P_{ijk}(L,d) \cdot N_{(f_i + f_j - f_k)}(m,N) + P_{ijk}(L,d) \cdot N_{(2f_i - f_k)}(m,N),$$
(6)

где $N_{(f_i+f_j-f_k)}(m,N)$ и $N_{(2f_i-f_k)}(m,N)$ – количество комбинационных частот типа $f_i + f_j - f_k$ и $2f_i - f_k$, соответственно, попадающих в спектральный канал с индексом *m* при передаче *N* спектральных каналов; $P_{ijk}(L,d)$ – мощность комбинационных частот на расстоянии *L* от начала ввода мультиплексированного сигнала в оптоволокно [5]:

$$P_{ijk}(L,d) = \eta_{ijk} \cdot \left(\frac{d}{3}\gamma Z_{eff}\right)^2 \cdot P_i \cdot P_j \cdot P_k \exp(-\alpha L), \qquad (7)$$

где η_{ijk} – эффективность генерации комбинационных частот; d – коэффициент вырождения (d = 6 для комбинационных частот типа $f_i + f_j - f_k$ и d = 3 для комбинационных частот типа $2f_i - f_k$); γ – нелинейный коэффициент оптоволокна; $L_{eff} = (1 - \exp(-\alpha L))/\alpha$ – эффективная длина оптоволокна [3]; α – коэффициент затухания оптоволокна.

Эффективность генерации комбинационных частот определяется следующим выражением:

$$\eta_{ijk} = \frac{\alpha^2}{\alpha^2 + \Delta\beta^2} \left[1 + \frac{4 \cdot \exp(-\alpha L)\sin^2(\Delta\beta L/2)}{(1 - \exp(-\alpha L))^2} \right],\tag{8}$$

где $\Delta\beta$ – фазовое рассогласование между передаваемыми спектральными каналами.

На основе результатов, полученных в [4, 5], можно утверждать, что наибольшая концентрация нежелательных комбинационных частот ЧВС приходится на спектральные каналы, находящиеся в середине частотного диапазона размещения спектральных каналов, поэтому в дальнейшем будем оценивать влияние ЧВС для спектрального канала в середине диапазона размещения. Определить количество комбинационных частот типа $f_i + f_j - f_k$ и $2f_i - f_k$, попадающих в спектральный канал с индексом *m* при передаче *N* спектральных каналов, можно при помощи следующих выражений [5]:

$$N_{(f_i+f_j-f_k)}(m,N) = \frac{m}{2}(N-m+1) + \frac{1}{4}((N-3)^2 - 5) - \frac{1}{8}(1-(-1)^N)(-1)^{N+m});$$
(9)

$$N_{(2f_i - f_k)}(m, N) = \frac{1}{2} (N - 2 - \frac{1}{2} (1 - (-1)^N) (-1)^m).$$
⁽¹⁰⁾

Явления ФСМ, ФКМ, ЧВС и шум ASE являются независимыми друг от друга [6, 7]. Тогда суммарный *Q*-фактор может быть определён следующим образом [2]:

$$1/Q_{\Sigma}^{2} \approx 1/Q_{NL+II}^{2} + 1/Q_{ASE}^{2} + 1/Q_{FWM}^{2} , \qquad (11)$$

где $Q_{NL+II} - Q$ -фактор детектируемого сигнала из-за влияния ФСМ и ФКМ с учётом наличия межсимвольной интерференции [3]; $Q_{ASE} - Q$ -фактор детектируемого сигнала из-за влияния шума ASE [2]; $Q_{FWM} - Q$ -фактор детектируемого сигнала из-за влияния ЧВС.

4. Результаты моделирования

Компьютерное моделирование проводилось в среде моделирования MathCAD. Результаты моделирования представлены семейством графиков изменения Q-фактора детектируемого сигнала в зависимости от уровня суммарной мощности оптического излучения в оптоволокне при различных значениях количества спектральных каналов и частотного интервала между спектральными каналами. Частотный интервал между спектральными каналами и количество каналов выбраны таким образом, чтобы обеспечить возможность передачи группового сигнала с заданной скоростью (приводятся результаты для скорости сигналов 1.6 Тбит/с). Таким образом, была смоделирована работа системы передачи в двух случаях: передача N = 160 спектральных каналов со скоростью 10 Гбит/с, форматом кодирования NRZ и частотным интервалом между спектральными каналами 12.5 ГГц и передача N = 40 спектральных каналов со скоростью 40 Гбит/с, форматом кодирования NRZ и частотным интервалом между спектральными каналами 50 ГГц. Для указанного числа спектральных каналов было определено количество комбинационных частот, попадающих в полосу длин волн спектральных каналов, находящихся в середине частотного диапазона размещения спектральных каналов (соответствует наихудшему случаю). Результаты представлены в табл. 1

Таблица 1. Максимальное количество частот, попадающих в полосу частот одного спектрального канала при разном количестве спектральных каналов N

Количество	Количество комбинационных частот		
спектральных каналов	$N_{(f_i+f_j-f_k)}$	$N_{(2f_i - f_k)}$	
N = 40	551	19	
N = 160	9401	79	

Результаты компьютерного моделирования представлены на рис. 2 в виде графиков зависимости Q_{NL+II} (кривая 1), Q_{ASE} (кривая 2), Q_{FWM} (кривая 3), Q_{Σ} (кривая 4) от уровня суммарной мощности оптического излучения в оптоволокне при различных параметрах системы передачи: число спектральных каналов и скорость передачи.



Рис. 2. Зависимость Q_{NL+II} (кривая 1), Q_{ASE} (кривая 3), Q_{FWM} (кривая 2), Q_{Σ} (кривая 4) от уровня суммарной мощности оптического излучения в оптоволокне: а) скорость передачи 10 Гбит/с, N = 160, тип кода NRZ; б) скорость передачи 40 Гбит/с, N = 40, тип кода NRZ

Характер кривых изменения Q-фактора детектируемого сигнала из-за влияния явления ЧВС, показанных на рис. 2, можно объяснить зависимостью ЧВС от количества спектральных каналов и частотного интервала между спектральными каналами. Как следует из формул (9) и (10), а также табл. 1, количество спектральных каналов напрямую влияет на количество комбинационных частот, попадающих в полосу длин волн спектрального канала [5]. Отметим, что изменение частотного интервала между спектральными каналами оказывает влияние на фазовое рассогласование между взаимодействующими спектральными каналами. Из рис. 2 видно, что уменьшение частотного интервала между спектральными каналами каналами приводит к снижению Q_{FWM} , что связано с уменьшением фазового рассогласования между взаимодействующими спектральными каналами между взаимодействующими спектральными каналами между взаимодействующими спектральными каналами приводит к снижению Q_{FWM} , что связано с уменьшением фазового рассогласования между взаимодействующими спектральными каналами между взаимодействующими спектральными между взаимодействующими спектральными каналами.

Из рис. 2 видно, что влияние явлений ФСМ и ФКМ на помехоустойчивость детектируемого сигнала значительно зависит от уровня суммарной мощности оптического излучения в оптоволокне. Подобную зависимость можно объяснить тем, что явления ФСМ и ФКМ возникают вследствие эффекта Керра, при котором показатель преломления среды передачи зависит от интенсивности оптического излучения, внося фазовые искажения в передаваемый сигнал.

Как видно из рис. 2, в приведенных результатах при разных скоростях передачи не происходит изменения кривой помехоустойчивости детектируемого сигнала под воздействием шума ASE. Это объясняется тем, что в данной работе при изменении скорости передачи также происходит изменение количества спектральных каналов. В результате в формуле (2) почти пропорционально изменяется не только знаменатель (увеличивается/уменьшается частотная полоса сигнала), но и числитель (увеличивается/уменьшается уровень мощности канальных сигналов).

При малом уровне суммарной мощности вводимого в оптоволокно оптического излучения основным фактором, ограничивающим помехоустойчивость детектируемого сигнала, является шум ASE. Причиной тому является слабая зависимость дисперсии фототока из-за шума ASE от мощности детектируемого сигнала [2]. Следует отметить, что в случае расположения спектральных каналов с частотным интервалом около 12.5 ГГц при малых уровнях суммарной мощности вводимого оптического излучения (до 5 мВт) явление ЧВС оказывает большее влияние на помехоустойчивость детектируемого сигнала по сравнению с явлениями ФСМ и ФКМ. Однако, ввиду сильной зависимости явлений ФСМ и ФКМ от уровня мощности сигнала, с увеличением уровня суммарной мощности оптического излучения, вводимого в оптоволокно, основным фактором, ограничивающим помехоустойчивость детектируемого сигнала, становится влияние явлений ФСМ и ФКМ. Также на графиках присутствует выраженная область с пиковым значением величины результирующего Q-фактора. Данная область соответствует оптимальному уровню суммарной мощности излучения в волокне с точки зрения помехоустойчивости сигналов к совместному воздействию нелинейных явлений (ФСМ, ФКМ, и ЧВС) и шума ASE. Также из рисунков видно, что при изменении скорости передачи и количества спектральных каналов N область пикового значения остается на том же значении уровня суммарной мощности излучения. Подобное поведение Q-фактора можно объяснить перераспределением уровня мощности сигналов при изменении количества спектральных каналов N.

5. Заключение

Реализована модель оценки влияния явлений ФСМ, ФКМ, ЧВС и шума ASE на помехоустойчивость сигнала в многоканальной ВОСП-DWDM. Критерием оценки изменения помехоустойчивости принят *Q*-фактор. На основе результатов, полученных в процессе компьютерного моделирования, можно сделать следующие выводы:

1. При малых уровнях мощности оптического излучения, вводимого в оптоволокно, основным фактором, ограничивающим помехоустойчивость сигнала, является шум ASE. В случае увеличения уровня суммарной мощности свыше 5 мВт при передаче 160 спектральных каналов со скоростью 10 Гбит/с в каждом спектральном канале и свыше 5 мВт при передаче 40 спектральных каналов со скоростью 40 Гбит/с в каждом спектральном канале основным фактором, ограничивающим помехоустойчивость детектируемого сигнала, становится влияние явлений ФСМ и ФКМ.

2. В случае использования частотного интервала между спектральными каналами, равного 12.5 ГГц, при уровнях мощности суммарного оптического излучения до 5 мВт явление ЧВС оказывает большее влияние на помехоустойчивость детектируемого сигнала по сравнению с влиянием явлений ФСМ и ФКМ. Дальнейшее увеличение уровня суммарной мощности оптического излучения выше 5 мВт приводит к преобладанию влияния явлений ФСМ и ФКМ на помехоустойчивость детектируемого сигнала.

3. Изменение величины частотного интервала между спектральными каналами приводит к снижению влияния явления ЧВС на помехоустойчивость детектируемого сигнала.

Следует отметить, что полученные результаты показывают вклад в снижение помехоустойчивости детектируемого сигнала только явлений ФСМ, ФКМ, ЧВС и шума ASE без учёта других линейных и нелинейных явлений. Данные результаты могут быть использованы при проектировании ВОСП-DWDM с целью оценки теоретически ожидаемого вклада явлений ФСМ, ФКМ, ЧВС и шума ASE в снижение помехоустойчивости детектируемого сигнала.

Литература

- 1. ITU-T Recommendation G.694.1 (02/2012). Spectral grids for WDM applications: DWDM frequency grid. URL: https://www.itu.int/rec/T-REC-G.694.1-201202-S/en (дата обращения: 01.04.2022).
- Sychuk A., Vardanyan V. Effect of Phase Noise on Pulsed Signals in a DWDM Optical Transmission System Using Coherent Detection // Proc. XV International Scientific-Technical Conference on Actual Problems of Electronic Instrument Engineering (APEIE), 2021. P. 351–354. URL: https://ieeexplore.ieee.org/document/9647626 (дата обращения: 01.04.2022).
- 3. Сычук А. Н., Варданян В. А. Искажение импульсных сигналов в многоканальных системах передачи с когерентным детектированием, вызванное явлениями фазовой самомоду-

ляции и фазовой кросс-модуляции в оптическом волокне // Т-Сотт: Телекоммуникации и транспорт. 2020. Т. 14, № 1. С. 4–12.

- 4. Варданян В. А. Влияние помех от четырехволнового смешения на спектрально разделенные каналы в пассивных оптических сетях доступа // Автометрия. 2017. Т. 53, № 1. С. 63–72.
- 5. *Варданян В. А.* Исследование распределения продуктов четырехволнового смешивания в ВОСП с ЧРК // Вестник СибГУТИ. 2016. № 2. С. 78–84.
- 6. Листвин В. Н., Трещиков В. Н. DWDM системы. М.: Наука, 2013. 261 с.
- 7. Agrawal G. P. Lightwave technology: telecommunication system. Hoboken, USA: Wiley-Interscience, 2005. 461 p.
- 8. *Shneider T.* Nonlinear optic in telecommunications. Springer-Verlag Berlin Heidelberg, 2004. 415 p.

Статья поступила в редакцию 27.04.2022; переработанный вариант – 29.05.2022.

Сычук Анатолий Николаевич

аспирант СибГУТИ, e-mail: anatolijsychuk@mail.ru.

Варданян Вардгес Андраникович

д.т.н., профессор кафедры фотоники в телекоммуникациях СибГУТИ (630102, Новосибирск, ул. Кирова, 86), e-mail: vardgesvardanyan@mail.ru.

DWDM signals noise immunity evaluation in a nonlinear fiber-optic path

Anatolij Sychuk

Postgraduate student, Siberian State University of Telecommunications and Information Science (SibSUTIS, Novosibirsk, Russia), anatolijsychuk@mail.ru.

Vardges A. Vardanyan

Doctor of technical sciences, Professor, Siberian State University of Telecommunications and Information Science (SibSUTIS, Novosibirsk, Russia), vardgesvardanyan@mail.ru.

The total influence of self-phase modulation, phase cross-modulation and four-wave mixing effects and amplified spontaneous emission noise on transmitted pulsed signals noise immunity in a multichannel fiberoptic transmission system with spectral division multiplexing is investigated. The investigation is carried out under the conditions of transmitting system parameters variation: the total optical power level in optical fiber, the transmission speed and the number of spectral channels. The estimate criterion of transmitted pulsed signals noise immunity is *Q*-factor. The total Q-factor value dependence on the total optical power level in optical fiber for various transmission speed and the number of spectral channels is demonstrated.

Keywords: pulsed signal noise immunity, wavelength division multiplexing, cross-phase modulation effect, four-wave mixing effect, amplified spontaneous emission noise.

References

- 1. ITU-T Recommendation G.694.1 (02/2012). Spectral grids for WDM applications: DWDM frequency grid, available at: https://www.itu.int/rec/T-REC-G.694.1-201202-S/en (accessed 01.04.2022).
- 2. Sychuk A., Vardanyan V. Effect of Phase Noise on Pulsed Signals in a DWDM Optical Transmission System Using Coherent Detection. 2021 XV International Scientific-Technical Conference on Actual Problems of Electronic Instrument Engineering (APEIE), 2021, pp. 351-354 available at: https://ieeexplore.ieee.org/document/9647626 (accessed 01. 04. 2022).
- Sychuk A. N., Vardanyan V. A. Iskazhenie impul'snyh signalov v mnogokanal'nyh sistemah peredachi s kogerentnym detektirovaniem, vyzvannoe yavleniyami fazovoj samomodulyacii i fazovoj kross - modulyacii v opticheskom volokne [Pulsed signals distortions in multichannel coherent detection transmission systems caused by self-phase modulation and cross-phase modulation in optical fiber]. *T-Comm: Telekommunikacii i transport*, 2020, vol. 14, no. 1, pp. 4-12.
- 4. Vardanyan V. A. Vliyanie pomekh ot chetyrekhvolnovogo smesheniya na spektral'no razdelennye kanaly v passivnyh opticheskih setyah dostupa [Four-wave mixing interference influence on wavelength division channels in passive optical access networks]. *Avtometriya*, 2017, vol. 53, no. 1, pp. 63-72.
- 5. Vardanyan V. A. Issledovanie raspredeleniya produktov chetyrekhvolnovogo smeshivaniya v VOSP s CHRK [Product distribution research of four-wave mixing in fiber optical transmission systems with frequency directed channels]. *Vestnik SibGUTI*, 2016, no. 2, pp. 78-84.
- 6. Listvin V. N., Treshchikov V. N. *DWDM sistemy* [Dense wavelength division multiplexing systems]. Moscow, Nauka, 2013, 261 p.
- 7. Agrawal G. P. *Lightwave technology: telecommunication system*. Hoboken, USA, Wiley-Interscience, 2005, p. 461.
- 8. Shneider T. Nonlinear optic in telecommunications. Springer-Verlag Berlin Heidelberg, 2004. p. 415.

Повышение точности позиционирования в навигационной системе автономного необитаемого подводного аппарата

В. Г. Арсентьев, Г. И. Криволапов¹

Предложен модифицированный алгоритм позиционирования навигационной системы автономного необитаемого подводного аппарата, повышающий точность навигации в условиях малых наклонных расстояний между объектами позиционирования.

Ключевые слова: автономный необитаемый подводный аппарат, навигационная система, алгоритм позиционирования.

1. Введение

В публикациях [1, 2] рассмотрен алгоритм позиционирования, ориентированный на минимально возможное количество приёмных трактов пеленгации и позволяющий определять координаты источника навигационного сигнала в трёхмерном пространстве.

В статье [3] даются оценки точностей позиционирования источника навигационного сигнала при использовании в навигационной системе фазового пеленгатора на базе четырёхэлементных пеленгационных антенн диаметрально-ортогональной (ДО) и пирамидальной геометрий.

В работе [4] предложена гидроакустическая навигационная система автономного необитаемого подводного аппарата (АНПА), использующая четырёхэлементные пеленгационные антенны ДО и пирамидальной геометрий, с размещением (рис. 1), позволяющим перекрывать переднюю, нижнюю и верхнюю полусферы приёма навигационного сигнала от маяка-ответчика объекта позиционирования.



1 – бортовые пеленгационные антенны диаметрально-ортогональной геометрии;
 2 – носовая пеленгационная антенна пирамидальной геометрии;
 3 – передающие антенны запросного сигнала.

Рис. 1. Размещение антенн навигационной системы АНПА

¹ Работа выполнена в рамках государственного задания № 071-03-2022-001.
С целью расширения диапазона рабочих частот навигационной системы в качестве навигационных сигналов маяков-ответчиков объектов позиционирования рекомендовано использовать амплитудно-модулированные (AM) сигналы с линейным детектированием в парциальных трактах приёма пеленгационных антенн АНПА [5].

Исходя из требований приемлемой технологической установки пеленгационных антенн в корпусе АНПА и характера изменения погрешностей оценок угловых координат источника навигационного сигнала по всем возможным направлениям выбранной полусферы приёма [3], предложены к применению: в качестве носовой антенны – четырёхэлементная антенна пирамидальной геометрии (форма правильной пирамиды), а в качестве бортовых антенн – четырёхэлементные антенны ДО-геометрии (форма правильного параллелепипеда) [2]. При этом следует обеспечить соответствующие геометрические параметры указанных антенн: для пирамидальной геометрии – размер рёбер основания пирамиды должен вдвое превышать её высоту, для ДО-геометрии – размер боковых рёбер параллелепипеда должен быть вдвое меньше размера рёбер основания.

Для указанных геометрических параметров пеленгационных антенн угловые координаты (пеленг φ и угол места θ) источника навигационного сигнала (передающей антенны) объекта позиционирования в системе координат пеленгационной антенны АНПА определяются на основе соответствующих тригонометрических соотношений.

Пеленгационная антенна пирамидальной геометрии:

$$\varphi = \begin{cases} \arccos\left(\frac{2\Delta\psi_{12} - \Delta\psi_{13} - \Delta\psi_{14}}{\sqrt{\left(2\Delta\psi_{12} - \Delta\psi_{13} - \Delta\psi_{14}\right)^{2} + 3\left(\Delta\psi_{13} - \Delta\psi_{14}\right)^{2}}}\right), \text{ если } 5(\Delta\psi_{13} - \Delta\psi_{14}) \ge 0 \\ \pi + \arccos\left(\frac{\Delta\psi_{13} + \Delta\psi_{14} - 2\Delta\psi_{12}}{\sqrt{\left(2\Delta\psi_{12} - \Delta\psi_{13} - \Delta\psi_{14}\right)^{2} + 3\left(\Delta\psi_{13} - \Delta\psi_{14}\right)^{2}}}\right), \text{ если } 5(\Delta\psi_{13} - \Delta\psi_{14}) < 0 \end{cases} ; (1) \\ = \begin{bmatrix} 0 \le \varphi < 2\pi \end{bmatrix}; \end{cases}$$

$$\theta = \arccos\left(\frac{1.1586(\Delta\psi_{12} + \Delta\psi_{13} + \Delta\psi_{14})}{\sqrt{(2\Delta\psi_{12} - \Delta\psi_{13} - \Delta\psi_{14})^2 + 3(\Delta\psi_{13} - \Delta\psi_{14})^2 + 1.3456(\Delta\psi_{12} + \Delta\psi_{13} + \Delta\psi_{14})^2}}\right) - \frac{\pi}{2};$$

$$[0 \le \theta \le \frac{\pi}{2}].$$
(2)

Пеленгационная антенна ДО-геометрии:

$$\varphi = \begin{cases}
 \operatorname{arccos}\left(\frac{-\Delta\psi_{12}}{\sqrt{\Delta\psi_{12}^{2} + (\Delta\psi_{13} - \Delta\psi_{14})^{2}}}\right), \ \operatorname{если} \Delta\psi_{13} - \Delta\psi_{14} \ge 0 \\
 \pi + \operatorname{arccos}\left(\frac{\Delta\psi_{12}}{\sqrt{\Delta\psi_{12}^{2} + (\Delta\psi_{13} - \Delta\psi_{14})^{2}}}\right), \ \operatorname{если} \Delta\psi_{13} - \Delta\psi_{14} < 0
 \end{cases};$$

$$\theta = \operatorname{arccos}\left(\frac{\Delta\psi_{12} - \Delta\psi_{13} - \Delta\psi_{14}}{\sqrt{\Delta\psi_{12}^{2} + (\Delta\psi_{13} - \Delta\psi_{14})^{2} + (\Delta\psi_{12} - \Delta\psi_{13} - \Delta\psi_{14})^{2}}}\right) - \frac{\pi}{2};$$

$$[0 \le \theta \le \frac{\pi}{2}].$$
(3)

Переменные $\Delta \psi_{12}$, $\Delta \psi_{13}$, $\Delta \psi_{14}$ в формулах (1) ÷ (4) являются инструментально измеренными разностями фаз колебаний демодулированного [5] навигационного АМ-сигнала маякаответчика объекта позиционирования для трёх пар парциальных трактов приёма пеленгационной антенны АНПА. Указанные соотношения являются частными случаями общего алгоритма позиционирования [1, 2] применительно к пеленгационным антеннам рассматриваемых геометрий.

Дальность *D* и наклонное расстояние *R* между пеленгационной антенной АНПА и передающей антенной маяка-ответчика объекта позиционирования могут быть оценены с использованием следующих зависимостей:

$$D = \frac{\left|h_1 - h_2\right|}{\operatorname{tg}\theta}, \ R = \frac{\left|h_1 - h_2\right|}{\sin\theta}.$$
(5)

Здесь h_1 – глубина погружения АНПА, полученная от датчика глубины; h_2 – глубина погружения передающей антенны объекта позиционирования, переданная на АНПА в информационной составляющей навигационного сигнала.

Погрешности оценок угловых координат φ и θ источника навигационного сигнала обусловлены погрешностями инструментального измерения разностей фаз $\Delta \psi_{12}$, $\Delta \psi_{13}$, $\Delta \psi_{14}$ и погрешностями самого алгоритма позиционирования – алгоритмическими погрешностями. Последние являются результатом используемых ограничений при математическом синтезе алгоритма [1, 2] и зависят от местоположения объекта позиционирования, геометрии и размеров пеленгационных антенн АНПА.

Для наклонных расстояний между АНПА и объектом позиционирования, превышающих 20 м, влияние на точность пеленгования алгоритмических погрешностей на порядок меньше влияния погрешностей инструментального измерения разностей фаз. Соизмеримое влияние на точность позиционирования алгоритмические погрешности оказывают при малых наклонных расстояниях (R < 10 м) между АНПА и объектом позиционирования, что нежелательно при решении задачи позиционирования с повышенной точностью, например, заведения АНПА в торпедный отсек подводной лодки, в причальные устройства подводной платформы техобслуживания [6] или надводного роботизированного аппарата.

2. Модифицированный алгоритм позиционирования

В основу предлагаемого модифицированного алгоритма позиционирования положено то обстоятельство, что зависимости погрешностей позиционирования от угловых координат объекта позиционирования, имеющие место при использовании соотношений (1) \div (4), являются гладкими функциями, у которых малые изменения аргумента приводят к малым изменениям функции. Поэтому в предлагаемом алгоритме позиционирования расчётные угловые координаты передающей антенны объекта позиционирования, использувания h_1 АНПА и глубине погружения h_2 передающей антенны объекта позиционирования, используя соотношения (1) \div (5), вычисляются сферические координаты передающей антенны объекта позиционирования, используя соотношения (1) \div (5), вычисляются сферические координаты передающей антенны объекта позиционирования, используя соотношения. Полученные сферические координаты передающей антенны объекта позиционирования, используя соотношения (1) \div (5), вычисляются сферические координаты передающей антенны объекта позиционирования, используя соотношения поредающей антенны объекта позиционирования, используя соотношения (1) \div (5), вычисляются сферические координаты передающей антенны объекта позиционирования, используя соотношения поредающей антенные сферические координаты передающей антенны объекта позиционирования, используя соотношения поредающей антенны объекта позиционирования, используемые затем для определения вспомогательных угловых координат передающей антенны объекта позиционирования с составляющими только алгоритмической погрешности, что позволяет оценить приближённые алгоритмиче

ские погрешности позиционирования и скорректировать результаты первоначально вычисленных угловых координат передающей антенны объекта позиционирования с уменьшенными алгоритмическими погрешностями.

На этой основе модифицированный алгоритм позиционирования предполагает в своей реализации следующий порядок действий и функционально связанных вычислительных операций.

1. В процессе позиционирования измеряются с достаточно малыми погрешностями: разности фаз $\Delta \psi_{12}$, $\Delta \psi_{13}$, $\Delta \psi_{14}$ колебаний демодулированного навигационного AM-сигнала маяка-ответчика объекта позиционирования для трёх пар гидрофонов пеленгационной антенны соответствующей полусферы приёма АНПА (см. рис. 1), глубина погружения h_1 АНПА и глубина погружения h_2 передающей антенны объекта позиционирования.

2. На основе соотношений (1) \div (5) в зависимости от полусферы приёма навигационного сигнала вычисляются сферические координаты передающей антенны объекта позиционирования: пеленг φ , угол места θ и наклонное расстояние R.

3. Вычисляются декартовы координаты x, y, z передающей антенны объекта позиционирования в системе координат пеленгационной антенны соответствующей полусферы приёма АНПА:

$$x = R\cos\theta\cos\varphi \; ; \; y = R\cos\theta\sin\varphi \; ; \; z = -R\sin\theta \; . \tag{6}$$

4. Для пеленгационных антенн АНПА с базовым установочным размером А (размер боковых рёбер параллелепипеда и высота пирамиды) в зависимости от полусферы приёма навигационного сигнала вычисляются расстояния S_1 , S_2 , S_3 , S_4 от каждого из приёмных гидрофонов до передающей антенны объекта позиционирования.

Антенна ДО-геометрии:

$$S_{1} = \sqrt{(A - x)^{2} + y^{2} + z^{2}}; S_{2} = \sqrt{(A + x)^{2} + y^{2} + z^{2}};$$

$$S_{3} = \sqrt{x^{2} + (A - y)^{2} + (A + z)^{2}}; S_{4} = \sqrt{x^{2} + (A + y)^{2} + (A + z)^{2}}.$$
(7)

Антенна пирамидальной геометрии:

$$S_{1} = \sqrt{x^{2} + y^{2} + (A + z)^{2}}; S_{2} = \sqrt{(1.155A - x)^{2} + y^{2} + z^{2}};$$

$$S_{3} = \sqrt{(0.58A + x)^{2} + (A - y)^{2} + z^{2}}; S_{4} = \sqrt{(0.58A + x)^{2} + (A + y)^{2} + z^{2}}.$$
(8)

5. Вычисляются вспомогательные угловые координаты φ_0 , θ_0 передающей антенны объекта позиционирования с составляющими только алгоритмической погрешности.

Антенна ДО-геометрии:

$$\varphi_{0} = \begin{cases} \arccos\left(\frac{S_{2} - S_{1}}{\sqrt{\left(S_{1} - S_{2}\right)^{2} + \left(S_{4} - S_{3}\right)^{2}}}\right), \text{ если } S_{4} - S_{3} \ge 0\\ \pi + \arccos\left(\frac{S_{1} - S_{2}}{\sqrt{\left(S_{1} - S_{2}\right)^{2} + \left(S_{4} - S_{3}\right)^{2}}}\right), \text{ если } S_{4} - S_{3} < 0 \end{cases} ;$$

$$[0 \le \varphi_{0} < 2\pi] ;$$

$$(9)$$

$$\theta_{0} = \arccos\left(\frac{S_{3} + S_{4} - S_{1} - S_{2}}{\sqrt{\left(S_{1} - S_{2}\right)^{2} + \left(S_{4} - S_{3}\right)^{2} + \left(S_{3} + S_{4} - S_{1} - S_{2}\right)^{2}}}\right) - \frac{\pi}{2};$$

$$\left[0 \le \theta_{0} \le \frac{\pi}{2}\right].$$

$$(10)$$

Антенна пирамидальной геометрии:

$$\varphi_{0} = \begin{cases}
 \operatorname{arccos}\left(\frac{S_{3} + S_{4} - 2S_{2}}{\sqrt{\left(S_{3} + S_{4} - 2S_{2}\right)^{2} + 3\left(S_{4} - S_{3}\right)^{2}}}\right), \text{ если } 5\left(S_{4} - S_{3}\right) \ge 0 \\
 \pi + \operatorname{arccos}\left(\frac{2S_{2} - S_{3} - S_{4}}{\sqrt{\left(S_{3} + S_{4} - 2S_{2}\right)^{2} + 3\left(S_{4} - S_{3}\right)^{2}}}\right), \text{ если } 5\left(S_{4} - S_{3}\right) < 0 \\
 \vdots \\
 10 \le \varphi_{0} < 2\pi]; \\
 \theta_{0} = \operatorname{arccos}\left(\frac{1.1586(3S_{1} - S_{2} - S_{3} - S_{4})}{\sqrt{\left(S_{3} + S_{4} - 2S_{2}\right)^{2} + 3\left(S_{4} - S_{3}\right)^{2} + 1.3456\left(3S_{1} - S_{2} - S_{3} - S_{4}\right)^{2}}}\right) - \frac{\pi}{2}; \\
 \left[0 \le \theta_{0} \le \frac{\pi}{2}\right].
 \end{aligned}$$
(11)

6. Вычисляются приближённые алгоритмические погрешности позиционирования $\Delta \varphi_A$, $\Delta \theta_A$:

$$\Delta \varphi_{\rm A} = \varphi_0 - \varphi \; ; \; \Delta \theta_{\rm A} = \theta_0 - \theta. \tag{13}$$

7. Вычисляются скорректированные угловые координаты $\varphi_{c\kappa}$, $\theta_{c\kappa}$ передающей антенны объекта позиционирования с уменьшенными алгоритмическими погрешностями:

$$\varphi_{c_{\rm K}} = \varphi - \Delta \varphi_{\rm A}; \ \theta_{c_{\rm K}} = \theta - \Delta \theta_{\rm A}. \tag{14}$$

В качестве показательного примера эффективности предложенного модифицированного алгоритма позиционирования на рис. 2 для пеленгационных антенн двух геометрий с базовым установочным размером A = 0.06 м показано расчётное уменьшение (в разах) максимальных алгоритмических погрешностей в зависимости от угла места позиционируемого объекта в области пеленгов $\varphi \in 0^{0}$... 360^{0} относительно курса АНПА при наклонных расстояниях до объекта R = 2 и 5 м, измеряемых с погрешностью ± 2 %, и погрешности инструментального измерения разностей фаз $\Delta \psi_{12}$, $\Delta \psi_{13}$, $\Delta \psi_{14}$ колебаний демодулированного АМ навигационного сигнала, равной $\pm 0.2^{0}$.

Представленные на рис. 2 зависимости показывают возможность значительного уменьшения алгоритмических погрешностей позиционирования в условиях малых наклонных расстояний между АНПА и объектом позиционирования, свидетельствуя о целесообразности применения предлагаемого алгоритма позиционирования в бортовой навигационной системе АНПА.

Кроме того, рассмотренный модифицированный алгоритм позиционирования оказывается востребованным на этапе предэксплуатационной градуировки антенн указанных геометрий фазового пеленгатора навигационной системы АНПА, выполняемой в лабораторных условиях по нестандартизованным методикам проверки функциональных элементов навигационного оборудования необитаемого подводного аппарата, которые относятся к категории малозатратных технологий и заслуживают самостоятельного рассмотрения.



Рис. 2. Уменьшение алгоритмической погрешности позиционирования

3. Заключение

Предложенный модифицированный алгоритм позиционирования может быть использован при решении задачи приведения автономных необитаемых подводных аппаратов к объектам причаливания различного вида и функционального назначения. Техническая реализация алгоритма не требует дополнительных аппаратурных затрат, так как выполняется на программном уровне. Алгоритм хорошо интегрируется в общую структуру навигационного комплекса автономного необитаемого подводного аппарата, не требуя значительных вычислительных ресурсов бортовой вычислительной системы. Возможность применения алгоритма при лабораторной оценке точностных характеристик фазовых пеленгаторов рассмотренной структуры указывает на его дополнительную практическую востребованность.

Литература

- 1. *Арсентьев В. Г., Криволапов Г. И.* Позиционирование объектов в гидроакустической навигационной системе с ультракороткой базой // Вестник СибГУТИ. 2018. № 4. С. 66–75.
- 2. Патент РФ 2709100, МПК G01S 1/72. Способ определения местоположения подводного объекта / В. Г. Арсентьев, Г. И. Криволапов, А. Е. Малашенко, Д. Д. Минаев. Заявка 2018122532, заявлено 19.06.2018, опубликовано 16.12.2019. Бюл. № 35.
- 3. *Арсентьев В. Г., Криволапов Г. И.* О влиянии геометрических параметров антенны на характеристики гидроакустического фазового пеленгатора // Вестник СибГУТИ. 2019. № 1. С. 92–101.
- 4. Арсентьев В. Г., Криволапов Г. И. Гидроакустическая навигационная система автономного необитаемого подводного аппарата // Сборник материалов РНТК «Современные проблемы телекоммуникаций», СибГУТИ, 2021. С. 159–168.
- 5. *Арсентьев В. Г., Криволапов Г. И.* Гидроакустический фазовый пеленгатор с амплитудномодулированным навигационным сигналом // Вестник СибГУТИ. 2021. № 2. С. 14–26.
- 6. *Арсентьев В. Г., Криволапов Г. И.* О характеристиках фазового пеленгатора гидроакустической системы приведения автономного необитаемого подводного аппарата // Вестник СибГУТИ. 2021. № 1. С. 23–35.

Статья поступила в редакцию 29.04.2022; переработанный вариант – 25.05.2022.

Арсентьев Виктор Георгиевич

к.т.н., ведущий научный сотрудник научно-технического центра специализированных информационных систем СибГУТИ (630008, Новосибирск, ул. Бориса Богаткова, 51), тел. (383) 2-693-938, e-mail: viktor.arsentev.51@mail.ru.

Криволапов Геннадий Илларионович

к.т.н., доцент, заведующий лабораторией, руководитель научно-технического центра специализированных информационных систем СибГУТИ, тел. (383) 2-693-942, e-mail: krivolapov@sibsutis.ru.

Improving the positioning accuracy in the navigation system of an autonomous uninhabited underwater vehicle

Viktor G. Arsent'ev

Candidate of technical sciences, Leading researcher, Siberian State University of Telecommunications and Information Science (SibSUTIS, Novosibirsk, Russia), viktor.arsentev.51@mail.ru.

Gennagy I. Krivolapov

Candidate of technical sciences, Head of laboratory, Siberian State University of Telecommunications and Information Science (SibSUTIS, Novosibirsk, Russia), krivolapov@sibsutis.ru.

A modified algorithm for positioning the navigation system of an autonomous uninhabited underwater vehicle is proposed. It increases the accuracy of navigation in conditions of small inclined distances between positioning objects.

Keywords: autonomous uninhabited underwater vehicle, navigation system, positioning algorithm.

References

- 1. Arsent'ev V. G., Krivolapov G. I. Pozicionirovanie ob"ektov v gidroakusticheskoj navigacionnoj sisteme s ul'trakorotkoj bazoj [Positioning of objects in a hydroacoustic navigation system with an ultrashort base]. *Vestnik SibGUTI*, 2018, no. 4, pp. 66-75.
- Arsent'ev V. G., Krivolapov G. I., Malashenko A.E., Minaev D.D. Sposob opredeleniya mestopolozheniya podvodnogo ob"ekta [Method of determining the location of an underwater object]. Patent RF, no. 2709100, IPC 2 G01S 1/72, 2019.
- 3. Arsent'ev V. G., Krivolapov G. I. O vliyanii geometricheskih parametrov antenny na harakteristiki gidroakusticheskogo fazovogo pelengatora [On the influence of the geometrical parameters of the antenna on the characteristics of the hydroacoustic phase direction finder]. *Vestnik SibGUTI*, 2019, no. 1, pp. 92-101.
- 4. Arsent'ev V. G., Krivolapov G. I. Gidroakusticheskaya navigacionnaya sistema avtonomnogo neobitaemogo podvodnogo apparata [Hydroacoustic navigation system of an autonomous uninhabited underwater vehicle]. *Sbornik materialov Rossijskoj nauchno-tekhnicheskoj konferencii: Sovremennye problemy telekommunikacij*, Novosibirsk, 2021, pp. 159-168.
- 5. Arsent'ev V. G., Krivolapov G. I. Gidroakusticheskij fazovyj pelengator s amplitudno-modulirovannym navigacionnym signalom [Hydroacoustic phase direction finder with amplitude-modulated navigation signal]. *Vestnik SibGUTI*, 2021, no. 2, pp. 14-26.
- 6. Arsent'ev V. G., Krivolapov G. I. O harakteristikah fazovogo pelengatora gidroakusticheskoj sistemy privedeniya avtonomnogo neobitaemogo podvodnogo apparata [On the characteristics of the phase direction finder of the hydroacoustic guidance system of an autonomous uninhabited underwater vehicle]. *Vestnik SibGUTI*, 2021, no. 1, pp. 23-35.

Алгоритм сбалансированной интерполяции пикселей при выделении контуров объектов на границах изображения

Д. В. Заерко, Н. Л. Боброва

Работа алгоритмов выделения контуров объектов, использующих операцию двумерной свертки, на границах растровых изображений всецело зависит от местоположения преобразуемого пикселя в пиксельной матрице. Проблема связана с необходимостью использования в самом алгоритме свертки дополнительного набора пикселей, находящихся вне пиксельной матрицы обрабатываемого изображения [1, 2, 3]. В данной работе предлагается к рассмотрению алгоритм сбалансированной интерполяции пикселей на границах изображения в дополнение к алгоритмам на основе статистических средних величин [4, 5]. Рассматриваются различные наборы коэффициентов, находящихся при определенной схеме расположения пикселей и участвующих в алгоритме интерполяции. Проводится сравнительный анализ результатов работы алгоритма сбалансированной интерполяции с результатами, полученными при использовании методов среднеарифметического взвешенного и моды по значениям дисперсии. Также приводится пример использования нового алгоритма для решения задачи выделения контуров объектов на границах растрового полутонового изображения. Описанный алгоритм прост для восприятия и использования, легко интегрируется с другими алгоритмами и не требует значительных вычислительных мощностей.

Ключевые слова: полутоновые изображения, пиксельная матрица, выделение контуров объектов, операция двумерной свертки, нелинейные операторы, интерполированные пиксели, алгоритм сбалансированной интерполяции.

1. Введение

С развитием технологий визуальных средств, улучшением каналов передачи данных, повышением возможностей физического хранения и обработки этих данных – роль технических ограничений в процессе распознавания постепенно снижается и основными ограничениями теперь выступают алгоритмические проблемы и их несовершенство. Неверно выбранный алгоритм или неэффективно работающий сведёт на нет все преимущества технических новшеств. С другой стороны, эффективный алгоритм позволит извлечь как можно больше полезной информации с имеющихся данных, открывая таким образом новые аспекты данных и многократно увеличивая возможности уже существующей технической инфраструктуры без дополнительных финансовых вложений. Большинство систем распознавания не уделяют должного внимания эффективности работы алгоритмов и опираются на широко распространенный, однако спорный принцип: «больше данных лучше, чем хороший алгоритм» [6]. Например: нет возможности воспроизвести повторно интересующий нас момент или событие для фиксации (снимки небесных тел, которые изменяют свое расположение постоянно). При обработке уже имеющегося набора данных, которые нельзя дополнить новыми данными (медицинская томография, криминалистика). Использование старой инфраструктуры одновременно с жесткими критериями по времени обработки и к результату. Например: в сфере

аэрофотосъемки или спутниковой съемки, где соединение нескольких изображений требует дополнительного снимка, чтобы убрать линии соединений снимков.

Во всех вышеперечисленных случаях, как и в других, где проводится обработка или распознавание изображений, применяются фильтры – специальные рода матрицы. Вид матрицы (фильтра) зависит от поставленной задачи, однако алгоритмов воздействия этих матриц на оригинальное изображения не так много, и одна из них – это операция свертки или двумерной свертки для матрицы. Как известно [1, 2, 3], алгоритм двумерной свертки имеет существенный недостаток при обработке граничных пикселей растрового изображения - алгоритм требует дополнительного набора пикселей вне изображения для корректной обработки последних пикселей изображения. Проблема актуальна для целого спектра алгоритмов, базирующихся на операции двумерной свертки, однако популярные и примитивные методы ее решения не дают приемлемого результата при конкатенации нескольких изображений высокого разрешения [4, 5, 7, 8]. Алгоритмы нелинейной фильтрации не являются исключением, используя для выделения контуров объектов на растровых изображениях принципы работы двумерной свертки: алгоритмы Собеля, Прюитта, Шарра. Прямая зависимость между результатами использования алгоритмов выделения контуров объектов на краях изображения от интерполированных пикселей указывает на необходимость выбора оптимального алгоритма с точки зрения значения дисперсии пикселей до и после интерполяции, а также вычислительных операций. В ходе сравнительного анализа методов интерполирования пикселей вне краев изображения статистическими методами [4, 5] с целью дальнейшего их использования в алгоритмах нелинейной фильтрации возникла гипотеза о неравномерном влиянии близлежащих пикселей на интерполируемый пиксель и прямую зависимость степени влияния от их расположения. Задача выделения контуров объектов на растровых изображениях относится к проблеме сегментации изображений. Сегментация изображения – процесс разделения цифрового изображения на сегменты по некоторому признаку, основная цель которого – упростить восприятие информации на изображении. Она является важной частью процесса распознания.

Использование верно интерполируемых пикселей позволит выполнить операцию двумерной свертки без пиксельных искажений в алгоритмах выделения контуров объектов и в результате приведёт к четкости границ объектов. Существующие методы интерполяции пикселей вне изображения решают проблему заполнения пикселями, не учитывая особенности этих пикселей – их яркость, сочетаемость с близлежащими и т.д. Первоначальная попытка проводить интерполяцию пикселей вне изображения по близлежащим с помощью общеизвестных статистических методов среднего выявила ряд практических вопросов [5]:

– Одинаково ли влияют близлежащие пиксели на интерполируемый пиксель? Если нет, то какие из них должны участвовать в интерполяции или каков вид схемы их расположения относительно интерполируемого пикселя?

– По какому критерию оценить влияние пикселей, а также какие весовые коэффициенты сопоставить для пикселей, участвующих в интерполяции?

– Имеет ли преимущество сбалансированный алгоритм интерполяции, построенный на основе весовых коэффициентов, перед другими методами? В частности, перед методом среднеарифметического взвешенного или моды, которые дали наилучшие результаты [5].

Объект исследования в статье – пиксельная матрица полутонового изображения. Предмет изучения – оценка степени влияния близлежащих пикселей на значения интерполируемых пикселей вне изображения и описание ее в виде нормированного набора весовых коэффициентов. Основная цель – описание алгоритма сбалансированной интерполяции пикселей вне изображения и подбор нормированных коэффициентов для него для практического использования при решении задачи выделения контуров объектов на растровых изображениях. Рассмотрим подробнее описанные проблемы и ответим на вопросы, поставленные выше. Проведем построение алгоритма сбалансированной интерполяции пикселей вне изображения в контексте задачи выделения контуров объектов на растровом изображения.

2. Выбор схемы для алгоритма сбалансированной интерполяции пикселей вне краев изображения

Определение схемы расположения пикселей в пиксельной матрице оригинального изображения играет важную роль в процессе интерполяции пикселей. Ранее в алгоритмах интерполяции использовались пиксели на расстоянии двух единичных шагов относительно интерполируемого пикселя. Пример такой схемы представлен на рис. 1.

$$\begin{array}{cccc} \nearrow a_1 & \rightarrow a_4 & \rightarrow a_7 \\ a^+ \rightarrow a_2 & \rightarrow a_5 & \rightarrow a_8 \\ \searrow a_3 & \rightarrow a_6 & \rightarrow a_9 \end{array}$$

Рис. 1. Схема элементов пиксельной матрицы при интерполяции пикселя вне изображения в центральной части крайнего левого столбца

Такое расположение и число элементов было обусловлено числом пикселей, используемых в матрице фильтра для операции двумерной свертки. Размер матрицы свертки составлял 3×3 . Это гарантировало, что число пикселей относительно невелико, и это не приводило к учету «чужих» значений, которые затрагивали бы уже контур нескольких объектов. Однако даже такое небольшое число элементов матрицы может затрагивать пиксели, не относящиеся непосредственно к контуру объекта на краях изображения и выступающие фоном. В свою очередь, фон может влиять на значение интерполируемого пикселя, исходя из своего веса среди общего числа анализируемых пикселей.

При оценке весовых коэффициентов для выбранной схемы возникает вопрос: какой алгоритм выбрать для интерполяции на «начальном» этапе? Если отталкиваться от предположения, что каждый пиксель имеет одинаковое влияние на интерполируемый, то возможно использовать общеизвестный метод среднего арифметического для интерполяции пикселей. Затем произвести вычисление коэффициента множественной корреляции для каждого из используемых в схеме пикселей относительно интерполируемого. Используя корреляционную оценку пикселей, выбрать весовые коэффициенты для пикселей, где коэффициенты нормализованы и в сумме равны единице $\sum_{k} b_k = 1, \sum_{p} c_p = 1$. Так будет получена приблизительная

оценка весовых коэффициентов. Полученные весовые коэффициенты будут учитываться в алгоритме сбалансированной интерполяции. Их дальнейшую подгонку можно проводить, изменяя нормированные весовые коэффициенты и наблюдая за изменением дисперсий пиксельных значений до и после включения интерполируемого пикселя в схеме. В случае пикселей с малыми весовыми коэффициентами выполняется изменение вида схемы (рис 2). Корреляционный анализ позволит оценить влияние выбора схем расположения близлежащих пикселей на интерполируемый пиксель и выбрать схему с наибольшим влиянием.

$$\begin{array}{cccc}
\uparrow a_1 \\
\nearrow a_2 & \nearrow a_6 \\
a^+ & \rightarrow a_3 & \rightarrow a_7 \\
& \searrow a_4 & \searrow a_8 \\
& \downarrow a_5
\end{array}$$

Рис. 2. Улучшенная схема расположения элементов пиксельной матрицы при интерполяции пикселя вне изображения в центральной части крайнего левого столбца

Проведем корреляционную оценку пикселей по улучшенной схеме на рис. 2 и используем в качестве источника данных пиксели с левой стороны растрового изображения (рис. 3) и его полутонового представления (рис. 4).



Рис. 3. Схема кварталов Барселоны

Рис. 4. Полутоновый вид схемы кварталов Барселоны

Корреляционную оценку выполним, используя возможности программных пакетов Wolframe Mathematica 9 и MS Excel. Пусть a^+ – интерполируемый пиксель вне изображения, выступающий в качестве зависимой переменной, $a_1, ..., a_8$ – смежные пиксели к интерполируемому пикселю a^+ , используемые как объясняющие переменные, располагаемые согласно схеме на рис. 2. Результаты корреляции приведем в табл. 1. Коэффициенты корреляционной матрицы в табл. 1 характеризуют тесноту связи между пикселями или, точнее, кодами полутонов пикселей.

	a^+	a_1	<i>a</i> ₂	<i>a</i> ₃	a_4	<i>a</i> ₅	a_6	<i>a</i> ₇	a_8
<i>a</i> ⁺	1								
a_1	0.81	1							
<i>a</i> ₂	0.85	0.82	1						
<i>a</i> ₃	0.95	0.67	0.82	1					
<i>a</i> ₄	0.83	0.62	0.67	0.82	1				
<i>a</i> ₅	0.64	0.54	0.62	0.67	0.82	1			
<i>a</i> ₆	0.82	0.75	0.83	0.71	0.62	0.58	1		
<i>a</i> ₇	0.78	0.64	0.75	0.83	0.71	0.62	0.88	1	
a_8	0.71	0.59	0.64	0.75	0.84	0.71	0.76	0.87	1

Таблица 1. Результаты корреляции пикселей по улучшенной схеме на рис. 2

Оценивая результаты столбца a^+ и анализируя корреляционные коэффициенты, указанные в строках столбца, делаем следующие выводы: 1. Существует корреляционная связь между анализируемыми пикселями из схемы на рис. 2 с интерполируемым пикселем вне изображения, так как все коэффициенты корреляции имеют значения 0.65 и более.

2. Анализ изменений значений коэффициентов от расположения в схеме на рис. 2 указывает на уменьшение влияния по мере удаления от интерполируемого пикселя.

3. Описанная на рис. 2 схема позволяет выполнять интерполяцию для пикселей вне изображения.

3. Формульные выражения весовых коэффициентов алгоритма сбалансированной интерполяции пикселей вне изображения

На основе коэффициентов корреляции и схемы расположения пикселей, используемых в процессе интерполяции, приведем строгую математическую формулировку задачи.

Рассмотрим алгоритм сбалансированной интерполяции пикселей вне изображения на базе оригинальной пиксельной матрицы $A = \{a_{i,j}\}, i = \overline{1, m}; j = \overline{1, m}$ растрового изображения и дальнейшее его применение для выделения контуров объектов. Введем обозначения и представим основные формульные выражения, справедливые для вышеуказанного алгоритма. Пусть число строк и столбцов матрицы $A^+ = \{a_{i,j}^+\}$ с интерполируемыми пикселями равны $i = \overline{0, n+1}, j = \overline{0, m+1}$ относительно размера $i = \overline{1, n}; j = \overline{1, m}$ оригинальной матрицы A растрового изображения. Оригинальная матрица составляет ядро матрицы A^+ и отличается лишь числом строк и столбцов, обрамляющих оригинальную, т.е. по строке и столбцу для всех сторон изображения: $a_{i,j}^+ = a_{i,j}, i = \overline{1, n}; j = \overline{1, j}$. Для упрощения восприятия матрицы A^+ и ее элементов $\{a_{i,j}^+\}$ дополнительные обозначения, кроме вышеперечисленных, вводиться не будут.

Пиксели вне изображения, требующие интерполяции, условно можно сгруппировать по их расположению относительно пиксельной матрицы. Опишем группы.

1. К «угловым» пикселям относятся первый и последний пиксели в интерполируемой строке или столбце, т.е. $a_{0,0}^+; a_{0,m+1}^+; a_{n+1,0}^+; a_{n+1,m+1}^+$.

2. «Предугловые» – пиксели, смежные к угловым пикселям и расположенные в интерполируемых строках и столбцах на удалении не более одного пикселя к угловому, т.е. $a_{1,0}^+$; $a_{1,m+1}^+$; $a_{0,m}^+$; $a_{n+1,1}^+$; $a_{n,0}^+$; $a_{n,m+1}^+$; $a_{n+1,m}^+$.

3. К «центральным» относятся все остальные пиксели, расположенные в интерполируемых строках и столбцах между предугловыми.

Определение конкретного набора весовых коэффициентов $b_1, b_2; c_1, c_2, c_3$ проводится путем их подбора с учетом оценки по наименьшей дисперсии и с соблюдением условия нормировки для каждой из групп коэффициентов, т.е. сумма весовых коэффициентов при интерполяции пикселя равна единице. Для интерполяции угловых и «предугловых» пикселей справедливо:

$$\begin{aligned} a_{0,0}^{+} &= a_{1,1}; a_{0,m+1}^{+} = a_{1,m}; a_{n+1,0}^{+} = a_{n,1}; a_{n+1,m+1}^{+} = a_{n,m}; \\ a_{0,1}^{+} &= a_{0,0}b_{1} + a_{1,1}b_{1} + a_{0,2}b_{1} + a_{1,0}b_{2} + a_{1,2}b_{2}, a_{0,2} = a_{1,2}, a_{1,0} = a_{1,1}; \\ a_{1,0}^{+} &= a_{0,0}b_{1} + a_{1,1}b_{1} + a_{2,0}b_{1} + a_{0,1}b_{2} + a_{2,1}b_{2}, a_{2,0} = a_{2,1}, a_{0,1} = a_{1,1}; \\ a_{1,m+1}^{+} &= a_{1,m}b_{1} + a_{2,m+1}b_{1} + a_{0,m+1}b_{1} + a_{0,m}b_{2} + a_{2,m}b_{2}, a_{2,m+1} = a_{1,m}, a_{0,m} = a_{1,m}; \\ a_{0,m}^{+} &= a_{1,m}b_{1} + a_{0,m+1}b_{1} + a_{0,m-1}b_{1} + a_{1,m-1}b_{2} + a_{1,m+1}b_{2}, a_{0,m-1} = a_{1,m-1}, a_{1,m+1} = a_{1,m}; \end{aligned}$$

$$\begin{aligned} a_{n+1,1}^+ &= a_{n+1,0}b_1 + a_{n,1}b_1 + a_{n+1,2}b_1 + a_{n,0}b_2 + a_{n,2}b_2, a_{n+1,2} = a_{n,1}, a_{n,0} = a_{n,1}; \\ a_{n,0}^+ &= a_{n+1,0}b_1 + a_{n-1,0}b_1 + a_{n,1}b_1 + a_{n-1,1}b_2 + a_{n+1,1}b_2, a_{n-1,0} = a_{n-1,1}, a_{n+1,1} = a_{n-1,1}; \\ a_{n,m+1}^+ &= a_{n+1,m+1}b_1 + a_{n,m}b_1 + a_{n+1,m-1}b_1 + a_{n,m-1}b_2 + a_{n-1,m+1}b_2, a_{n+1,m-1} = a_{n,m-1}, a_{n-1,m+1} = a_{n-1,m}. \\ \text{Для интерпояции центральных пикселей воспользуемся выражениями:} \\ a_{n+1,m}^+ &= a_{n,m}b_1 + a_{n+1,m+1}b_1 + a_{n-1,m+1}b_1 + a_{n-1,m}b_2 + a_{n+1,m}b_2, a_{n-1,m+1} = a_{n-1,m}, a_{n+1,m} = a_{n,m}; \\ a_{i,j}^+ &= c_1a_{i,j+1} + c_2a_{i-1,j+1} + c_2a_{i+1,j+1} + c_3a_{i,j+1} + c_3a_{i+2,j+1} + c_3a_{i-1,j+2} + c_3a_{i,j+2} + c_3a_{i+1,j+2}; i = \overline{3,n-2}; j = 0; \\ a_{i,j}^+ &= c_1a_{i,j-1} + c_2a_{i+1,j-1} + c_2a_{i,j-1} + c_3a_{i+2,j-1} + c_3a_{i-2,j-1} + c_3a_{i-1,j-2} + c_3a_{i,j-2} + c_3a_{i+1,j-2}; i = \overline{3,n-2}; j = m + 1; \\ a_{i,j}^+ &= c_1a_{i-1,j} + c_2a_{i-1,j-1} + c_2a_{i-1,j+1} + c_3a_{i-1,j+2} + c_3a_{i-1,j-1} + c_3a_{i-2,j-1} + c_3a_{i-2,j-1}$$

+
$$c_3a_{i+2,j-1}+c_3a_{i+2,j+1}; i=0; j=\overline{3,m-2}$$
.
Теперь необходимо провести подбор значений коэффициентов для «предугловых» b_1, b_2
и центральных c_1, c_2, c_3 пикселей. Оценка справедливости используемых весовых коэффици-
ентов будет проводиться по значениям дисперсии до и после включения интерполируемого

4. Оценка весовых коэффициентов. Пример использования алгоритма

В качестве примера тестовых данных для дальнейшей оценки будут использоваться полутоновые значения пиксельной матрицы по изображению на рис. 4. Исходя из того, что интерес для анализа представляют только интерполируемые и граничные пиксели полутоновой матрицы изображения, и ввиду большой размерности пиксельной матрицы представим данные в графическом виде. Чтобы избежать загромождения излишними данными на рисунках и для сохранения всей полноты информации, достаточной для аргументации выводов, можно рассмотреть одну из сторон анализируемого изображения, например, левую (левый столбец пиксельной матрицы). Выбор стороны изображения произвольный. Аналогичные выводы можно получить, анализируя другие стороны.

Основные цели теста:

1. Провести интерполяцию пикселей вне изображения, используя алгоритм сбалансированной интерполяции с заданными весовыми коэффициентами, а также методами среднеарифметического взвешенного и моды.

2. Вычислить значения дисперсий интерполируемых и граничных смежных к ним кодов полутонов растрового изображения до и после включения в оценку интерполируемых значений.

3. Оценить используемые коэффициенты путем сравнения значений интерполируемых по сбалансированному алгоритму пикселей вне изображения с граничными смежными пикселями оригинального изображения. Сравнить дисперсии интерполируемых и смежных к ним кодов полутонов. Сравнить различные методы интерполяции по пиксельным значениям и по значениям их дисперсий.

4. В случае положительной оценки весовых коэффициентов для алгоритма сбалансированной интерполяции применить алгоритмы нелинейной фильтрации (Собеля, Прюитта,

пикселя.

Шарра) по интерполированным значениям. Сравнить результаты применения алгоритмов выделения контуров объектов по пикселям, интерполированных различными методами.

Пусть $b_1 = 0.25$; $b_2 = 0.125$ и $c_1 = 0.1$; $c_2 = 0.15$; $c_3 = 0.06$. Построим графики кодов полутонов для граничных смежных пикселей изображения и интерполированных пикселей вне изображения по алгоритму сбалансированной интерполяции, а также их дисперсии.

Из анализа графиков на рис. 5 видно, что значения кодов полутонов, полученные при помощи алгоритма сбалансированного интерполирования (красным цветом), близки к значениям пикселей оригинального изображения (черный цвет). Близкие, но не одинаковые, так как строятся на основе оценки множества пикселей. Такой принцип гарантирует, что в случае единичных пиксельных ошибок в кодах полутона интерполированное значение не будет значительно искажено. В большинстве случаев черный график совпадает или находится над красным – алгоритм сбалансированной интерполяции имеет свойство незначительно преуменьшать значения по отношению к граничным пикселям оригинального изображения. На рис. 7 и 9, где сравниваются коды полутонов, интерполированные по различным методам, красные графики (алгоритм сбалансированной интерполяции) совпадают с черными (мода и среднеарифметический взвешенный метод) на основных интервалах. На некоторых интервалах алгоритм сбалансированной интерполяции выстраивает скорректированные и меньшие значения. Происходит корректировка скачков, которые описывались ранее в [5]. Оценка средних значений дисперсий после учета пикселей без интерполяции Dorg, т.е. копии граничных пикселей оригинального изображения с использованием интерполяции пикселей методами среднеарифметического взвешенного Davg и моды Dmod в сравнении с результатами интерполяции сбалансированным алгоритмом *D*balanse имеют значения:

$$D_{balanse} = 331; D_{avg} = 428; D_{mod} = 654; D_{org} = 631.$$

Значения дисперсий для алгоритма сбалансированной интерполяции значительно меньше, чем при использовании других методов. Рост значений дисперсии после добавления интерполированного значения не наблюдался.

Улучшенная схема на рис. 2, используемая для интерполяции пикселей вне изображения с набором предложенных весовых коэффициентов, может быть использована в процессе интерполяции пикселей и показала лучшие результаты в сравнении с другими методами интерполяции.



Рис. 5. Коды полутонов для интерполированных сбалансированным алгоритмом и граничных пикселей изображения

 $\begin{array}{c}
D \\
5000 \\
4000 \\
3000 \\
2000 \\
1000 \\
0 \\
200 \\
400 \\
600 \\
800 \\
x
\end{array}$

Рис. 6. Дисперсия значений кодов полутонов для интерполированных сбалансированным алгоритмом и граничных пикселей изображения



Рис. 8. Дисперсия значений кодов полутонов для интерполированных по моде и сбалансированным алгоритмом пикселей



Рис. 9. Коды полутонов для интерполированных среднеарифметическим взвешенным методом и сбалансированным алгоритмом пикселей



Рис. 10. Дисперсия кодов полутонов для интерполированных среднеарифметическим взвешенным методом и сбалансированным алгоритмом пикселей

Следующим шагом в оценке алгоритма сбалансированной интерполяции будет использование интерполированных пикселей в алгоритмах выделения контуров объектов, которые, как упоминалось ранее, основаны на принципах работы операции двумерной свертки и требуют интерполяции пикселей вне границ оригинального изображения. Применяя операторы Собеля, Прюитта, Шарра, сравним результаты их применения для граничных смежных пикселей оригинального изображения, а также для пикселей, интерполированных методами среднеарифметического взвешенного, моды и алгоритмом сбалансированной интерполяции. Графики откликов по методу Собеля представлены на рис. 11 – 13.

R xРис. 11. Значения откликов R по интерполированным сбалансированным алгоритмом

и граничным пикселям изображения



Рис. 12. Значения откликов для интерполированных модой пикселей и сбалансированным алгоритмом



Рис. 13. Значения откликов R для интерполированных среднеарифметическим методом и сбалансированным алгоритмом пикселей

Графики откликов при использовании фильтра Шарра (рис. 14 – 16):



Рис. 14. Значения откликов R по интерполированным сбалансированным алгоритмом пикселям и граничным пикселям изображения



Рис. 16. Значения откликов R для интерполированных среднеарифметическим методом и сбалансированным алгоритмом пикселей



 x

Рис. 17. Значения откликов R по интерполированным сбалансированным алгоритмом



Рис. 18. Значения откликов для интерполированных модой и сбалансированным алгоритмом пикселей



Рис. 19. Значения откликов R для интерполированных среднеарифметическим методом и сбалансированным алгоритмом пикселей

R

Оценивая в целом графики откликов, для всех трех методов выделения контуров объектов прослеживается общая тенденция: графики, построенные с учетом интерполированных пикселей по алгоритму сбалансированной интерполяции, на большинстве интервалов строк совпадают с графиками других методов интерполяции. Имеющиеся различия касаются отдельных фрагментов строк, на которых алгоритм сбалансированной интерполяции выдает меньшие значения, чем при других методах интерполяции. Иными словами, в закономерностях по откликам прослеживаются схожие тенденции, так же как в графиках кодов полутонов. Это еще раз указывает на важность выбранного метода интерполяции пикселей вне изображения при решении задач алгоритмами на основе матриц двумерной свертки.

Приведенный пример не только носил иллюстрационный характер рассуждений при оценке нового алгоритма, но и имел практическую направленность. Решение задачи выделения контуров объектов на краях изображения целиком зависит от интерполяции пикселей вне изображения и выполняется в контексте выделения контуров объектов городской инфраструктуры. Перепад описывает разграничение между объектами на изображении или, оперируя терминами прикладной задачи, разграничение улицами домов и прочих объектов городской инфраструктуры. При детальном изучении графиков, увеличивая их размер, становится ясно, что увеличение значения отклика указывает на наличие границ контура некоторого объекта. В противном случае при относительной однородности значений и отсутствии явно выраженных скачков поведение значений укажет на промежутки между объектами (улицы и т.д.).

5. Заключение

Схема работы на основе операций двумерной свертки широко распространена среди прикладных алгоритмов, особенно в сфере сегментации. Проблема выхода алгоритма свертки за пределы обрабатываемой пиксельной матрицы существует и оказывает существенное воздействие на последующие этапы процесса обработки, в частности, на этап выделения контуров объектов на краях растрового изображения, что было продемонстрировано в примере статьи. В статье была доказана справедливость предположения о различии степени влияния на интерполируемые пиксели в зависимости от расположения смежных пикселей. Предположение было подтверждено как теоретически, опираясь на оценку результатов корреляционного анализа, так и на практике – по эмпирическому анализу графиков. На основе доказанного предположения была представлена улучшенная схема для алгоритма сбалансированной интерполяции и подобраны его весовые коэффициенты. Представленный алгоритм сбалансированной интерполяции пикселей дополняет уже существующие и используемые примитивными алгоритмами. Как показал анализ дисперсий и интерполированных значений, алгоритм позволяет получить лучшие результаты для дальнейшей обработки с несущественным увеличением сложности в отношении к иным алгоритмам. В то же время алгоритм открыт для улучшения в плане подбора более точного значения весовых коэффициентов.

Литература

- 1. *Хириман И. И, Уиддер Д. В.* Преобразования типа свертки. М.: Издательство иностранной литературы, 1958. 312 с.
- 2. Брейсуэлл Р. Н. Преобразование Хартли: пер с англ. М.: Мир, 1990. 175 с.
- 3. Dettmers T. Understanding convolution in deep learning // TD Blog. URL: https://timdettmers.com/2015/03/26/convolution-deep-learning (дата обращения: 20.03.2022).
- 4. Заерко Д. В., Липницкий В. А. Алгоритмическая проблема выделения контуров объектов на растровых изображениях // Технические средства защиты информации: тезисы докла-

дов XVII Белорусско-российской научно-технической конференции, Минск, 6 июня 2021 г.С. 45.

- 5. Заерко Д. В., Липницкий В. А. Анализ методов определения граничных пикселей полутонового изображения при операции двумерной свертки. // Новости науки и технологии. 2021. № 2. С. 43–52.
- 6. Хендрик Б., Джозеф Р., Марк Ф. Машинное обучение. СПб.: Питер, 2017. 336 с.
- 7. Заерко Д. В., Липницкий В. А. Алгоритм весового определения граничных пикселей // Системный анализ и прикладная информатика. 2022. № 4. С. 59–62.
- 8. *Bailey D. G.* Image border management for FPGA based filters // Proc. Sixth IEEE International Symposium on Electronic Design, Test and Application. Queenstown, New Zealand, 17–19 Jan 2011. P. 144–149.

Статья поступила в редакцию 03.05.2022; переработанный вариант – 23.06.2022.

Заерко Денис Владимирович

аспирант кафедры информатики БГУИР, e-mail: zaerko1991@gmail.com.

Боброва Наталия Леонидовна

к.т.н., доцент кафедры информатики БГУИР (Республика Беларусь, Минск, ул. Платонова, 39, корп. 5, каб. 401a), e-mail: bobrova@bsuir.by.

Balanced pixel interpolation algorithm when identifying the object's contours on the borders of the image

Denis V. Zaerko

Postgraduate student, Informatics department of Belarusian State University of Informatics and Radioelectronics (BSUIR, Minks, Republic Belarus), zaerko1991@gmial.com.

Natalia L. Bobrova

Ph. D of technical sciences, associate Professor of the Informatics Department of Belarusian State University of Informatics and Radioelectronics (BSUIR, Minks, Republic Belarus), bobrova@bsuir.by.

The operation of algorithms for identifying object's contours using two-dimensional convolution operation on the boundaries of raster images depends entirely on the location of the transformed pixel in the pixel matrix. The problem is related to the need for the convolution algorithm itself to use an additional set of pixels located outside the pixel matrix of the processed image [1, 2, 3]. In this paper, an algorithm for balanced interpolation of pixels at image boundaries in addition to algorithms based on statistical averages [4, 5] is proposed. Various sets of coefficients which are located at a certain pixel layout and participate in the interpolation algorithm is considered. A comparative analysis of the results of the balanced interpolation algorithm with the results using the arithmetic mean methods is carried out. Also, an example of using a new algorithm to solve the problem of the object's contours identification on the borders of a raster halftone image is presented. The described algorithm is easy to understand, use and integrate with other algorithms and it does not require significant computing power.

Keywords: halftone images, pixel matrix, object contour selection, two-dimensional convolution operation, non-linear operators, interpolated pixels, balanced interpolation algorithm.

References

- 1. Hirschman I. I., Widder D. V. *The convolution transform*. Princeton, N.J, Princeton University Press, 1955, 268 p.
- 2. Bracewell Ronald N. The Hartley Transform. Oxford University Press, New York, NY United States. 1986.
- 3. Dettmers T. Understanding convolution in deep learning *TD Blog.* available at: https://timdettmers.com/2015/03/26/convolution-deep-learning (accessed: 20.03.2022).
- 4. Zaerko D. V., Lipnitski. V. A. Algoritmicheskaja problema vydelenija konturov ob#ektov na rastrovyh izobrazhenijah [Algorithmic problem of selecting the contours of objects on bitmaps]. Theses of 17th RNTK «Technical means of information protection» Conference, Minsk, Republic Belarus, 6 June, 2021, p. 45.
- 5. Zaerko D V., Lipnitski. V. A. Analiz metodov opredeleniya granichnyh pikselej polutonovogo izobrazheniya pri operacii dvumernoj svertki [Analysis of methods and algorithms for determining boundary pixels of a half-tone image in a two-dimensional conversion operation]. News of science and technologies. 2021. no 2. pp 43-52.
- 6. Henrick B, Joseph W. R, Mark Fetherolf. Real Word Machine Learning, Manning Publications, 2016. 234 p.
- 7. Zaerko D. V., Lipnitski V. A. *Algoritm vesovogo opredelenija granichnyh pikselej* [Weighted determination algorithm of boundary pixels]. *System analysis and applied information science*. 2020. no. 4. pp. 59-62.
- 8. Bailey D. G. Image border management for FPGA based filters. 2011 Sixth IEEE International Symposium on Electronic Design, Test and Application. Queenstown, New Zealand, 17-19 Jan 2011, pp 144-149.

Верификация модели интегральной катушки индуктивности для СВЧ LC-фильтров в Si- и SiGe-системах на кристалле

В. В. Ерохин

В работе описывается модель проводника, которая может быть использована для построения катушек индуктивности различных конфигураций топологий в любых Si и SiGe технологических процессах. Для верификации модели произведены экспериментальные образцы тестовых катушек индуктивности в технологическом процессе SiGe БиКМОП 130 нм. Результаты испытаний микросхем показали, что характеристики опытных образцов находятся в диапазоне смоделированных значений с учетом технологического разброса. Предложенная модель имеет большую сходимость с характеристиками опытных образцов, чем 3D-моделирование топологии, при этом скорость моделирования эквивалентной схемы может на порядки превышать скорость 3D-моделирования. Точность модели обеспечивается учетом скин-эффекта и краевых эффектов в диэлектрике и в подложке. Использование эквивалентной схемы скин-эффекта позволяет производить симуляцию модели в Cadence Spectre Simulator, а также создавать модели CB4 LC-фильтров.

Ключевые слова: модель проводника, Si, SiGe, модель катушки индуктивности, скинэффект, краевой эффект.

1. Введение

Бурное развитие сферы беспилотных автомобилей и летательных аппаратов формирует требования к устройствам их управления и связи. Возрастают требования к габаритным размерам, потребляемой мощности и диапазону работы частот. Решением является использование устройств типа «система-на-кристалле» (СнК), где на одной подложке формируются все сложно-функциональные (СФ) аналоговые и цифровые блоки. Технологические процессы GaAs имеют малую степень интеграции (максимальное количество транзисторов в чипе измеряется десятками), поэтому они не подходят для создания сложных цифровых блоков. Si и SiGe технологические процессы имеют значительно большую степень интеграции и позволяют размещать в одном чипе десятки и сотни тысяч транзисторов, поэтому именно эти технологии подходят для разработки современных устройств типа СнК.

В широкополосных СВЧ-приемниках и передатчиках, которые используются в БПЛА, из-за плотного расположения каналов связи требуются LC-фильтры с крутизной спада АЧХ до 300 дБ/дек, следовательно, требуется использовать большое количество катушек индуктивности. Достоверность моделей катушек напрямую влияет на характеристики фильтров и на функционирование устройства в целом.

При проектировании интегральных устройств с катушками индуктивности у разработчика может возникать ряд трудностей и проблем:

- могут отсутствовать модели катушек индуктивности;
- модель может быть некорректная или иметь ошибки;
- могут отсутствовать правила экстрагирования индуктивностей.

Одним из решений является 3D-моделирование катушек индуктивности или пассивных устройств в целом. Данный метод требует дополнительного специализированного САПР, при этом 3D-моделирование требуют больших вычислительных мощностей и занимают значительно больше времени (на несколько порядков), чем моделирование эквивалентных схем или экстракций топологических описаний.

В данной работе будет предложена универсальная модель проводника для построения моделей катушек индуктивности различных конструкций.

Катушки индуктивности имеют и другие СФ-блоки СВЧ широкополосных СнК, такие как малошумящие усилители мощности, LC-генераторы и др. Поэтому данная работа будет полезна при проектировании СВЧ СнК в целом.

2. Модель проводника

На рис. 1 представлена П-модель интегрального проводника. L_s – индуктивность проводника, C_{OX} – емкость проводника к подложке, C_{sub} и R_{sub} – емкость и сопротивление подложки. Схема из элементов $L_1 - L_3$ и $R_1 - R_3$ является эквивалентной схемой сопротивления проводника с учетом скин-эффекта [1–5].

Рис. 1. П-модель интегрального проводника

2.1. Скин-эффект

С учетом скин-эффекта сопротивление проводника можно рассчитать, используя (1)-(3) [1-3]:

$$R_s(f) = \frac{\rho \cdot l}{w \cdot t_{eff}},\tag{1}$$

где l – длина, w – ширина, ρ – удельное сопротивление, t_{eff} – эффективная толщина проводника, рассчитываемая по формуле:

$$t_{eff}(f) = \delta \cdot \left(1 - e^{-\frac{t}{\delta}}\right),\tag{2}$$

где *t* – физическая толщина проводника, *δ* – глубина скин-эффекта:

$$\delta(f) = \sqrt{\frac{\rho}{\pi \cdot \mu \cdot \mu_0 \cdot f}},\tag{3}$$



где *µ*, *µ*⁰ – относительная магнитная проницаемость и магнитная проницаемость вакуума.

На большой частоте эффективная толщина проводника стремится к величине глубины скин-эффекта:

$$t_{eff}(f_{\max}) \approx \delta(f_{\max})$$

Рис. 2 характеризует модель скин-эффекта. При повышении частоты сигнала ток по проводнику распределяется неравномерно и преимущественно в поверхностном слое. $R_1 - R_4$ эквивалентны сопротивлению эффективной толщины проводника при различных частотах. От разбиения сечения проводника, т.е. от количества LR-сегментов зависит точность модели скин-эффекта. Модель 4-го уровня позволяет получить ошибку не более 2 %...5 %. Такую точность можно считать достаточной, так как ошибка модели скин-эффекта на порядок меньше, чем технологический разброс сопротивления проводников, который может достигать ± 20 %.



Рис. 2. Скин-эффект

Элементы схемы с частотно-зависимыми параметрами, такие как $R_s(f)$, невозможно реализовать напрямую в симуляторе Cadence Spectre, что является проблемой и в большинстве случаев при экстракции топологического описания скин-эффект абсолютно не учитывается. Представление скин-эффекта в виде эквивалентной схемы, состоящей из резисторов и катушек индуктивности, позволяет использовать ее в Spectre-моделях и производить частотные и временные моделирования в различных САПР, в том числе в Cadence Virtuoso.

Отношения между элементами эквивалентной схемы выражаются коэффициентами *K*_R и *K*_L [4–5].

$$R_1 = R_2 \cdot K_R = R_3 \cdot K_R^2 = R_4 \cdot K_R^3 , \qquad (4)$$

$$L_1 = L_2 \cdot K_L = L_3 \cdot {K_L}^2 \,. \tag{5}$$

Согласно формулам (1) – (3) сопротивление проводника с учетом скин-эффекта пропорционально \sqrt{f} . Для получения такой характеристики эквивалентной схемы коэффициенты выбираются в пределах: $K_R > 1$, $0 < K_L < 1$. С достаточной точностью можно принять, что $K_R = 1/K_L$.

Сопротивление эквивалентной схемы скин-эффекта при постоянном токе можно найти как параллельное сопротивление резисторов *R*₁ – *R*₄:

$$R_{s}(dc) = \frac{1}{\frac{1}{R_{1}} + \frac{K_{R}}{R_{1}} + \frac{K_{R}^{2}}{R_{1}} + \frac{K_{R}^{3}}{R_{1}}} = \frac{R_{1}}{K_{R}^{3} + K_{R}^{2} + K_{R} + 1}.$$
(6)

На высокой частоте сопротивление эквивалентной схемы стремиться к R_1 , из этого следует, что сопротивление R_1 должно быть не меньше сопротивления скин-эффекта при максимальной частоте $R_{\text{max}} = R_s(f_{\text{max}})$. В данной работе $f_{\text{max}} = 60$ ГГц.

$$\frac{R_{\max}}{R_s(dc)} \le K_R^3 + K_R^2 + K_R + 1.$$
(7)

$$\frac{t}{\delta(f_{\max})} \le K_R^{3} + K_R^{2} + K_R + 1.$$
(8)

Полное сопротивление эквивалентной схемы описывается выражениями (9)-(14):

$$Zs_{L3,R4}(j\omega) = R_4 + j\omega \cdot L_3, \qquad (9)$$

$$Zs_{R3,L3,R4}(j\omega) = \frac{R_3 \cdot Zs_{L3,R4}(j\omega)}{R_3 + Zs_{L3,R4}(j\omega)},$$
(10)

$$Zs_{L2,R3,L3,R4}(j\omega) = Zs_{R3,L3,R4}(j\omega) + j\omega \cdot L_2, \qquad (11)$$

$$Zs_{R2,L2,R3,L3,R4}(j\omega) = \frac{R_2 \cdot Zs_{L2,R3,L3,R4}(j\omega)}{R_2 + Zs_{L2,R3,L3,R4}(j\omega)},$$
(12)

$$Zs_{L1,R2,L2,R3,L3,R4}(j\omega) = Zs_{R2,L2,R3,L3,R4}(j\omega) + j\omega \cdot L_1,$$
(13)

$$Zs(j\omega) = \frac{R_1 \cdot Zs_{L1,R2,L2,R3,L3,R4}(j\omega)}{R_1 + Zs_{L1,R2,L2,R3,L3,R4}(j\omega)}.$$
(14)

При
$$K_R = 1/K_L$$
 полное сопротивление эквивалентной схемы будет равно:

$$Z_{S}(j\omega) = \left[\omega^{2}L_{1}^{2}R_{1}^{2}K_{R}^{4} \cdot \left(K_{R}^{4} + \frac{R_{1}}{R_{s}(dc)}\right) - R_{1}^{4} - j\omega L_{1}R_{1}K_{R} \cdot \left[R_{1}^{2}\left(K_{R}^{4} + K_{R}^{2} + \frac{R_{1}}{R_{s}(dc)}\right) - \omega^{2}L_{1}^{2}K_{R}^{8}\right]\right]/$$

$$/\left[\omega^{2}L_{1}^{2}R_{1}K_{R}^{4} \cdot \left(K_{R}^{5} + K_{R}^{4} + \frac{R_{1}}{R_{s}(dc)}\right) - \frac{R_{1}^{4}}{R_{s}(dc)} - j\omega L_{1}K_{R} \cdot \left[\frac{R_{1}^{3}K_{R}^{3}}{R_{s}(dc)} + R_{1}^{2}\left(K_{R}^{4} + K_{R}^{2} + \frac{R_{1}}{R_{s}(dc)}\right) - \omega^{2}L_{1}^{2}K_{R}^{8}\right]\right]/$$

$$(15)$$

Так как $L_3 > L_2 > L_1$, то на большой частоте при $t_{eff}(f) \approx \delta(f)$ сопротивление эквивалентной схемы будет примерно равно:

$$Zs(j\omega_{\max}) \approx \frac{\frac{R_1^2}{K_R} + R_1 \cdot j\omega_{\max} \cdot L_1}{R_1 + \frac{R_1}{K_R} + j\omega_{\max} \cdot L_1}.$$
(16)

Активное сопротивление при максимальной частоте:

$$Rs(f_{\max}) = \operatorname{Re}\left[Zs(j\omega_{\max})\right] \approx \frac{\frac{R_{1}^{3}}{K_{R}} \cdot \left(1 + \frac{1}{K_{R}}\right) + \omega_{\max}^{2} \cdot L_{1}^{2} \cdot R_{1}}{R_{1} \cdot \left(1 + \frac{1}{K_{R}}\right)^{2} + \omega_{\max}^{2} \cdot L_{1}^{2}}.$$
(17)

Из полученного выражения можно вычислить L₁:

$$L_{1} \approx \frac{R_{1} \cdot \left(1 + \frac{1}{K_{R}}\right)}{2 \cdot \pi \cdot f_{\max}} \cdot \sqrt{\frac{R_{s}(f_{\max}) - \frac{R_{1}}{K_{R} + 1}}{R_{1} - R_{s}(f_{\max})}}.$$
(18)

В [4, 5] приведено уравнение для нахождения *K*_{*L*}:

$$\frac{1}{K_L^2} + \frac{1}{K_L} \cdot \left(\frac{1}{K_R} + 1\right)^2 + \left(\frac{1}{K_R^2} + \frac{1}{K_R} + 1\right)^2 = 3.175 \cdot \left(\left(\frac{1}{K_R} + 1\right) \cdot \left(\frac{1}{K_R^2} + 1\right)\right)^2.$$
(19)

Решая квадратное уравнение, получим формулу:

$$K_{L} = \frac{\left(1 + \frac{1}{K_{R}}\right)^{2} + \sqrt{\left(1 + \frac{1}{K_{R}}\right)^{4} - 4 \cdot \left(\frac{1}{K_{R}^{2}} + \frac{1}{K_{R}} + 1\right)^{2} - 12.7 \cdot \left(\frac{1}{K_{R}^{3}} + \frac{1}{K_{R}^{2}} + \frac{1}{K_{R}} + 1\right)^{2}}{6.35 \cdot \left(\frac{1}{K_{R}^{3}} + \frac{1}{K_{R}^{2}} + \frac{1}{K_{R}} + 1\right)^{2} - 2 \cdot \left(\frac{1}{K_{R}^{2}} + \frac{1}{K_{R}} + 1\right)^{2}} \right)^{2} \cdot (20)$$

Алгоритм расчета эквивалентной схемы скин-эффекта:

- 1. Выбор минимального K_R из выражений (7), (8).
- 2. Расчет *R*₁ по формуле (6).
- 3. Расчет *L*₁ по формуле (18).
- 4. Расчет *K*^{*L*} по формуле (20).

5. Рассчитать ошибку эквивалентной схемы, и если она больше необходимой, то увеличить K_R и повторять пункты 1–5 до нахождения удовлетворительного значения ошибки.

Представленный алгоритм требует много итераций для получения ошибки не более 5 %. В данной работе для простоты вычислений было принято, что

$$K_R = \frac{1}{K_L} = 2$$

а оптимальную величину L_1 можно найти, используя формулу (15). Полученные значения индуктивности находятся в пределах $0.9 \times L_1 \dots 0.95 \times L_1$, от индуктивности, рассчитанной по формуле (18). Это позволило получить ошибку в частотном диапазоне 1...40 ГГц не более 5 % за одну итерацию.

2.2. Паразитные емкости

Емкость проводника к подложке можно оценить по формуле плоского конденсатора, одной обкладкой которого считать нижнюю плоскость проводника, а второй обкладкой – подложку:

$$C_{ox} = \frac{\varepsilon_r \cdot \varepsilon_0 \cdot w \cdot l}{2 \cdot t_{ox}} , \qquad (21)$$

где ε_0 – электрическая постоянная, ε_r – диэлектрическая проницаемость диэлектрика, t_{ox} – толщина диэлектрика между проводником и подложкой.

Если представить, что обкладки конденсатора имеют бесконечную длину (рис. 3а), то электрическое поле существует только между обкладками и все силовые линии электрического поля равны, параллельны и перпендикулярны пластинам. Классическая формула плоского конденсатора (21) описывает такой случай и справедлива, если *t*_{OX} << *l*, *w*. В реальности

обкладки имеют конечную длину, и силовые линии электрического поля замыкаются не только по самому короткому пути (рис. 3б), причем на краях обкладок напряженность электрического поля больше, чем в середине. Это называют краевыми эффектами плоского конденсатора.



Рис. 3. Емкость к подложке без учета краевых эффектов (а) и с учетом (б)

В литературе описано множество формул учета краевых эффектов конденсаторов [6, 7]. Рассмотрим некоторые из них. Палмер аппроксимировал краевые эффекты плоского конденсатора методом Шварца–Кристоффеля для обкладок конечной длины и бесконечно малой толщины:

$$C_{ox} = \frac{\varepsilon_r \cdot \varepsilon_0 \cdot w \cdot l}{2 \cdot t_{ox}} \cdot \left(1 + \frac{t_{ox}}{\pi \cdot l} + \frac{t_{ox}}{\pi \cdot l} \cdot \ln\left(\frac{2 \cdot \pi \cdot l}{t_{ox}}\right) \right).$$
(22)

Позже Чанг, используя метод Шварца–Кристоффеля, смог вывести формулы емкости с учетом краевых эффектов и толщины проводника:

$$C_{ox} = \frac{\varepsilon_r \cdot \varepsilon_0 \cdot w}{2 \cdot \pi} \cdot \ln\left(\frac{R_b}{R_a}\right),\tag{23}$$

$$R_a = \exp\left(-1 - \frac{\pi \cdot l}{2 \cdot t_{ox}} - a \tanh\left(\frac{1}{\sqrt{p}}\right) \cdot \frac{p+1}{\sqrt{p}} - \ln\left(\frac{p-1}{4p}\right)\right),\tag{24}$$

$$R_b = \eta + \frac{p+1}{2} \ln \Delta \,, \tag{25}$$

$$\eta = \sqrt{p} \cdot \left[\frac{\pi \cdot l}{2 \cdot t_{ox}} + \frac{p+1}{2 \cdot \sqrt{p}} \cdot \left(1 + \ln\left(\frac{4}{p-1}\right) \right) - 2 \cdot a \tanh\left(\frac{1}{\sqrt{p}}\right) \right], \tag{26}$$

$$\Delta = \max(\eta, p), \tag{27}$$

$$p = 2 \cdot B^2 - 1 + \sqrt{\left(2 \cdot B^2 - 1\right)^2 - 1}, \qquad (28)$$

$$B = 1 + \frac{t_M}{t_{ox}} \,. \tag{29}$$

Формула Палмера имеет расхождение с моделированием методом конечных элементов не более 1.3 %, а формулы Чанга – не более 1 % [6]. Технологический разброс диэлектрической проницаемости SiO₂ может достигать более 5 %, поэтому для расчета емкостей к подложке с учетом краевых эффектов целесообразно использовать формулу Палмера, которая

обеспечивает ошибку значительно меньше технологического разброса, при этом расчет значительно проще, чем по формулам Чанга.

2.3. Подложка

Емкость подложки *C_{sub}* и сопротивление *R_{sub}* подложки можно найти по формулам [8–10]:

$$C_{sub} = \frac{\varepsilon_{sub} \cdot \varepsilon_0 \cdot l}{4 \cdot F(w, t_{sub})},$$
(30)

$$F(w, t_{sub}) = \begin{cases} \frac{1}{2 \cdot \pi} \cdot \ln\left(\frac{8 \cdot t_{sub}}{w} + \frac{w}{4 \cdot t_{sub}}\right) & npu \quad \frac{t_{sub}}{w} > 1 \\ \frac{1}{\frac{w}{t_{sub}} + 2.42 - \frac{0.44 \cdot t_{sub}}{w} + \left(1 - \frac{t_{sub}}{w}\right)^6} & npu \quad \frac{t_{sub}}{w} < 1 \end{cases}$$
(31)
$$R_{sub} = \frac{4 \cdot F(w, t_{sub}) \cdot \rho_{sub}}{l \cdot \left[1 + \left(1 + \frac{10 \cdot t_{sub}}{w}\right)^{-\frac{1}{2}}\right]},$$
(32)

где t_{sub} , ε_{sub} , ρ_{sub} – толщина, диэлектрическая проницаемость и удельное сопротивление подложки.

Недостатком формул (30)–(32) является то, что здесь не учитывается краевой эффект. Это значит, что особенно при малых площадях проводников рассчитанные C_{sub} и R_{sub} будут значительно (в разы) отличаться от реальных.

Формула (33) получена методом изображений и учитывает краевой эффект поля, который состоит в том, что силовые линии электрического поля распространяются внутри подложки дальше физических краев проводника [11–13].

$$C_{sub} = \frac{\varepsilon_{sub} \cdot \varepsilon_0 \cdot w \cdot l}{2 \cdot t_{sub} + \sqrt{\frac{w \cdot l}{\pi}} - \sqrt{4 \cdot t_{sub}^2 + \frac{w \cdot l}{\pi}}}.$$
(33)

Тестовые измерения емкостей и сопротивлений подложки показали, что формула (33) имеет большую погрешность только при площади проводника более 40000 мкм² [13].

Сопротивление подложки находится по формуле времени диэлектрической релаксации т [7–10]:

$$R_{sub} = \frac{\tau}{C_{sub}} = \frac{\rho_{sub} \cdot \varepsilon_{sub} \cdot \varepsilon_0}{C_{sub}} \,. \tag{34}$$

2.4. Индуктивности

Собственную индуктивность прямого проводника прямоугольного сечения можно найти по известной формуле [13–14]:

$$L_{s} = \frac{\mu_{0} \cdot l}{2 \cdot \pi} \cdot \left[\ln\left(\frac{2 \cdot l}{w+t}\right) + 0,50049 + \frac{w+t}{3 \cdot l} \right].$$
(35)

3. Модель катушки индуктивности

На рис. 4 представлены микрофотография и топология экспериментального образца катушки индуктивности, выполненного в технологическом процессе SiGe БиКМОП 130 нм. Катушка индуктивности выполнена в медном слое металлизации, а контактные площадки и подводящие выводы – в алюминиевом слое.

На рисунке топологии экспериментального образца обозначены проводники и паразитные емкости между проводниками, выполненными в одном слое металлизации. Расшифровка обозначений представлена в табл. 1.



Рис. 4. Микрофотография (а) и топология (б) экспериментального образца катушки индуктивности

Проводники общего вывода (земли) дублируются во всех слоях металлизации с максимальным количеством переходных отверстий между ними, а также имеются 4 контактные площадки для подключения при измерении экспериментального образца, поэтому сопротивление проводника земли даже с учетом скин-эффекта будет стремиться к нулю и его величиной можно пренебречь.

Обозначение	Расшифровка
Z_{ind}	Проводник восьмиугольной катушки индуктивности
Zio Cu	Проводник вывода катушки в медном слое металлизации
Zio Al	Проводник вывода катушки в алюминиевом слое металлизации
R_{via}	Сопротивление массива переходных отверстий между слоями металлизаций
C_s	Межвитковая емкость катушки
C_{via}	Емкость между переходными отверстиями двух выводов катушки
Cio Cu	Емкость между выводами катушки в медном слое металлизации
C_{ind-0}	Емкость между проводником катушки индуктивности и проводником земли
C	Емкость между проводником вывода катушки в алюминиевом слое и провод-
Cio_Al-0	ником земли
C_{pad-0}	Емкость между контактной площадкой и проводником земли

Таблица 1. Обозначения элементов топологии

101

На рис. 5 представлена модель экспериментального образца катушки индуктивности. Для одновитковой катушки индуктивности C_s будет близок к нулю, поэтому этой емкостью можно пренебречь.

Можно заметить, что катушка индуктивности обозначена как *Z*_{ind}, однако нужно учесть, что индуктивность не будет равняться собственной индуктивности по формуле (35). Необходимо учесть взаимные индуктивности.



Рис. 5. Модель образца катушки индуктивности

Существует множество формул плоских катушек индуктивности [15, 16], такие как формула Уиллера, формулы, основанные на аппроксимации плотности тока, формулы, полученные в результате аппроксимации экспериментальных образцов. Недостаток этих формул заключается в том, что они узкоприменимы и имеют большие погрешности.

С большой точностью можно найти индуктивность катушек любой формы, используя метод разбиения катушки на прямые сегменты проводников и нахождения их собственных и взаимных индуктивностей [17–20].

Собственную индуктивность сегментов катушки можно найти по формуле (35), а взаимные индуктивности параллельных сегментов можно найти по формуле:

$$M_{seg} = \frac{\mu_0 \cdot Q \cdot l_{seg}}{2 \cdot \pi},\tag{36}$$

где l_{seg} – длина сегмента, Q – коэффициент взаимной индуктивности, который рассчитывается по формуле Гровера:

$$Q = \ln\left[\frac{l_{seg}}{GMD} + \sqrt{1 + \left(\frac{l_{seg}}{GMD}\right)^2}\right] - \sqrt{1 + \left(\frac{GMD}{l_{seg}}\right)^2} + \frac{GMD}{l_{seg}},$$
(37)

где *GMD* – среднегеометрическое расстояние между сегментами, которое рассчитывается по формуле:

$$GMD = \exp\left[\ln\left(D\right) - \frac{w^2}{12 \cdot D^2} - \frac{w^4}{60 \cdot D^4} - \frac{w^6}{168 \cdot D^6} - \frac{w^8}{360 \cdot D^8}\right],$$
(38)

где *D* – расстояние между центрами сегментов.

Взаимная индуктивность между сегментами одинаковой длины, соединенными одним концом, можно найти по формуле [17]:

$$M_{a_seg} = \frac{\mu_0}{\pi} \cdot l_{seg} \cdot \cos(\alpha) \cdot \operatorname{a} \tanh\left(\frac{l_{seg}}{l_{seg} + y}\right),\tag{39}$$

где у – расстояние между концами сегментов:

$$y = l_{seg} \cdot \sqrt{2 + 2 \cdot \cos(\alpha)}. \tag{40}$$

Взаимная индуктивность между сегментами, расположенными под углом 90°, равна нулю, а согласно формуле (39) положительные взаимные индуктивности сегментов, расположенных под углом 45°, равны отрицательным взаимным индуктивностям, расположенным под углом 135°. Тогда полную индуктивность одновитковой симметричной восьмиугольной катушки индуктивности можно найти по формуле:

$$L_{ind} = 8 \cdot L_{seg} - 16 \cdot M_{seg}, \tag{41}$$

где *L_{seg}* – собственная индуктивность одного сегмента.

Выводы катушки в медном слое параллельны и находятся достаточно близко, поэтому также следует учесть их взаимную индуктивность.

Сопротивление массива переходных отверстий можно найти по формуле:

$$R_{Via} = \frac{r_{Via}}{N_{Via}},\tag{42}$$

где *r*_{*Via*} – сопротивление одного переходного отверстия, *N*_{*Via*} – количество переходных отверстий в одном квадратном массиве:

$$\sqrt{N_{Via}} \le \frac{w - 2 \cdot c + b}{a + b},\tag{43}$$

где *а* – длина и ширина одного переходного отверстия, *b* – минимальное расстояние между отверстиями, *с* – минимальное расстояние от отверстия до края металла.

4. Результаты измерений и моделирований

Измерения экспериментального образца проводились GSG-зондами с помощью зондовой станции MPI TS200 и векторного анализатора спектра ZVA40, имеющего рабочий диапазон от 1 кГц до 40 ГГц. Перед испытаниями образца производилась калибровка измерительной установки на специализированной калибровочной плате компании MPI.

Моделирование эквивалентной схемы катушки индуктивности (рис. 5) и моделирования экстракции топологического описания производились с помощью ПО Cadence Virtuoso, 3Dмоделирование топологии катушки производилось в ПО Empire XPU.

Эффективные индуктивности и добротности экспериментального образца, модели и экстракции топологического описания были найдены по формулам (44) и (45).

$$L_{eff} = \frac{X_L}{2 \cdot \pi \cdot f} = \frac{\operatorname{Im}\left(\frac{1}{Y_{11}}\right)}{2 \cdot \pi \cdot f},\tag{44}$$

$$Q_{eff} = \frac{2 \cdot \pi \cdot f \cdot L_{eff}}{R_{eff}} = \frac{\operatorname{Im}\left(\frac{1}{Y_{11}}\right)}{\operatorname{Re}\left(\frac{1}{Y_{11}}\right)}.$$
(45)

На рис. 6 и 7 представлены графики эффективной индуктивности и добротности. Цветными линиями представлены измерения 7 экспериментальных образцов, случайным образом выбранных из партии, сплошными линиями представлены результаты симуляции предложенной модели катушки индуктивности, пунктирными линиями – результаты 3Dмоделирования топологии. Результаты моделирования экстракции топологического описания сильно отличаются от остальных результатов, поэтому они не приведены на графиках, а сведены в табл. 2.

Симуляции предложенной модели катушки и 3D-моделирования выполнены с учетом технологического разброса, то есть приведены графики в худшем и лучшем случаях по критерию максимальной добротности.



Таблица 2. Результаты моделирования экстракции топологического описания

Параметр	$L_{eff}(1 \Gamma \Gamma \mu), \pi \Gamma \mu$	f_r , ГГц	Q_{max}	$f(Q_{max}), \Gamma \Gamma$ ц
Значение	569	20.3	56.6	11.5

5. Обсуждение результатов

Полученные данные показывают, что предложенная модель имеет большое схождение с измерениями экспериментального образца. Результаты 3D-моделирования имеют величину добротности и резонансную частоту, близкие к реальному образцу, но частота максимальной добротности отличается в два раза и значение индуктивности на 15 % больше. Результаты моделирования экстракции топологического описания значительно отличаются от результатов испытаний: максимальная добротность в пять раз больше, частота максимальной добротности в два раза больше, резонансная частота в полтора раза меньше. В результате анализа файлов экстракции было выявлено, что правила экстракции, предоставляемые заводомизготовителем, предусматривают упрощенное экстрагирование подложки, вследствие чего результаты моделирования экстракции сильно отличаются от реальной микросхемы.

На графике эффективной добротности видно, что некоторые реальные образцы имеют максимальную добротность меньше, чем предложенная модель в худшем случае технологического разброса. Отчасти это может быть вызвано относительной простотой модели, так как учтены не все паразитные элементы, но более вероятно влияние контактного сопротивления зондов с контактными площадками. На рис. 7 приведены графики добротности экспериментальных образцов после нескольких контактов зондов с площадками, где видно, что измеренные добротности уменьшаются, следовательно, увеличивается контактное сопротивление. Как и говорилось ранее, перед измерениями была произведена калибровка на специализированной калибровочной пластине, однако материалом проводника является золото, которое значительно тверже алюминия, из которого изготовлены контактные площадки.

На рис. 8 представлены контактные площадки после измерений и их Z-профили, показывающие рельеф контактных площадок. Видно, что глубина проникновения зондов составляет 3 мкм, что является толщиной слоя металлизации, в котором располагаются контактные площадки. Это может вызывать контактные сопротивления, сопоставимые с активным сопротивлением катушки индуктивности.



Рис. 8. Контактные площадки после измерений

Для исключения контактного сопротивление между зондами и площадками следует производить калибровку непосредственно на чипе, в котором выполнено исследуемое устройство, однако это требует выделения большой площади кристалла, которая будет не задействована в работе устройства. На практике волновое сопротивление устройств составляет 50 Ом, и доли ома контактного сопротивления не оказывают серьезного влияния на результаты испытаний.

С учетом контактного сопротивления можно с уверенностью сказать, что характеристики экспериментального образца находятся в пределах лучшего и худшего случаев предложенной модели.

6. Заключение

Предложенная модель показывает сходимость с экспериментальным образцом лучше, чем 3D-моделирование и значительно лучше экстракции топологического описания. Это свидетельствует о том, что модели, представленные заводом-изготовителем, а также правила экстрагирования могут недостоверно отражать реальные характеристики интегральных индуктивностей.

Данная работа может являться основой для создания конфигурируемой Spectre-модели проводника и катушки индуктивности, так как здесь используется эквивалентная схема скинэффекта с частотно независимыми параметрами элементов.

В результате проведённой работы было выявлено, что для получения достоверной модели проводников и катушек индуктивности достаточно использовать П-модель проводника с учетом скин-эффекта, краевых эффектов и эффектов в подложке. Дальнейшее усложнение и уточнение модели приводит к увеличению точности, которую невозможно оценить ввиду того, что технологический разброс значительно больше.

Полученные данные актуальны при разработке интегральных СВЧ LC-фильтров, а также могут быть применены и при разработке других аналоговых СФ-блоков СВЧ СнК.

Литература

- 1. *Ruehli A. E., Antonini G., Jiang L.* Skin-Effect Model for Round Wires in PEEC // Proc. International Symposium on Electromagnetic Compatibility (EMC Europe). Rome, Italy, 2012. P. 6.
- Ren-Jia Chan, Jyh-Chyurn Guo. Analysis and Modeling of Skin and Proximity Effects for Millimeter-Wave Inductors Design in Nanoscale Si CMOS // Proc. 9th European Microwave Integrated Circuit Conference. Rome, Italy, 2014. P. 13–16.
- Rotella F. M., Blaschke V., Howard D. A Broad-Band Scalable Lumped-Element Inductor Model Using Analytic Expressions to Incorporate Skin Effect, Substrate Loss, and Proximity Effect // Digest. International Electron Devices Meeting, San Francisco, CA, USA, 2002. P. 471–474.
- Cao Yu, Groves R. A., Zamdmer N. D. Frequency-Independent Equivalent Circuit Model for On-chip Spiral Inductors // Proc. IEEE Custom Integrated Circuits Conference, Orlando, FL, USA, 2002. P. 217–220.
- Kim S., Neikirk D. P. Compact Equivalent Circuit Model for the Skin Effect // Proc. IEEE MTT-S International Microwave Symposium Digest, San Francisco, CA, USA, 1996. P. 1815–1818.
- Elsaadi M., Tayel M. B., Steenson D. P. An Empirical Formula of Fringing Field Capacitance for MEMS Tunable Capacitor Actuators // Proc. IEEE 1st International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA), Tripoli, Libya, 2021. P. 674–678.
- 7. *Ehud Yariv*. Edge corrections for parallel-plate capacitors // European Journal of Applied Mathematics, Cambridge, England, 2020. P. 226–241.
- 8. Ting Chen, Jin He, Pengwei Chen, Xuanyu Chen, Hao Wang, Sheng Chang, Qijun Huang. Accurate Modeling of Three-Port Center-tapped Octagonal Inductors for SPDT Switch Design in

0.13-µm BiCMOS // Proc. IEEE International Conference on Ubiquitous Wireless Broadband (ICUWB), Nanjing, China, 2016. P. 4.

- 9. Wei Yi Lim, M. Annamalai Arasu, M. Kumarasamy Raja. Modeling of Two Port Center-tapped to ground and Three Port scalable symmetrical inductor // Proc. International Symposium on Integrated Circuits (ISIC), Singapore, 2014. P. 540–543.
- 10. Chen J., Liou J. J. Improved and Physics-Based Model for Symmetrical Spiral Inductors // IEEE Transactions on Electron Devices. 2006. V. 53, Is. 6. P. 1300–1309.
- Goni A., del Pino J., Gonzalez B., Hernandez A. An Analytical Model of Electric Substrate Losses for Planar Spiral Inductors on Silicon // IEEE Transactions on Electron Devices. 2007. V. 54, Is. 3. P. 546–553.
- 12. Jeyaraman S., Vanukuru V., Nair D. R., Chakravorty A. A Substrate Model for On-Chip Tapered Spiral Inductors with Forward and Reverse Excitations // IEEE Transactions on Electron Devices. 2019. V. 66, Is. 1. P. 4.
- Jeyaraman S., Vanukuru V., Nair D. R., Chakravorty A. Compact Modeling of Series Stacked Tapered Spiral Inductors // Proc. IEEE 19th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems (SiRF), Orlando, FL, USA, 2019. P. 3.
- 14. Passos F., Fino M. H., and Moreno E. R. Fully Analytical Characterization of the Series Inductance of Tapered Integrated Inductors // International Journal of Electronics and Telecommunications, Poland, 2014. P. 73–77.
- 15. *Fino M. H.* Using an Integrated Inductor Model in Ques // Proc. 21st International Conference Mixed Design of Integrated Circuits and Systems (MIXDES), Lublin, Poland, 2014. P. 66–69.
- Zolog M., Pitica D., Pop O. Characterization of Spiral Planar Inductors Built on Printed Circuit Boards // Proc. 30th International Spring Seminar on Electronics Technology (ISSE), Cluj-Napoca, Romania, 2007. P. 308–313.
- 17. *Koutsoyannopoulos Y. K., Papananos Y.* Systematic Analysis and Modeling of Integrated Inductors and Transformers in RF IC Design // IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing. 2000. V. 47, Is. 8. P. 699–713.
- 18. Chen J., Liou J. J. On-Chip Spiral Inductors for RF Applications: An Overview // Journal of Semiconductor Technology and Science. 2004. P. 149–167.
- Shaltout A. H., Gregori S. Optimizing the Inductance Time-Constant Ratio of Polygonal Integrated Inductors // Proc. IEEE 61st International Midwest Symposium on Circuits and Systems (MWSCAS), Windsor, ON, Canada, 2018. P. 448–451.
- Heng-Ming Hsu, Jen-Zien Chang, Hung-Chi Chien. Coupling Effect of On-Chip Inductor with Variable Metal Width // IEEE Microwave and Wireless Components Letters. 2007. V. 17, Is. 7. P. 498–500.

Статья поступила в редакцию 7.06.2022; переработанный вариант – 24.06.2022.

Ерохин Виктор Валерьевич

научный сотрудник НИИРП НИЛ «Системы на кристалле» ОмГТУ (644050, Омск, просп. Мира, 11), e-mail: viktor_erohin@mail.ru.

Integrated Inductor Model Verification for Microwave LC-filters in Si and SiGe Systems on a Chip

Viktor V. Erokhin

Research scientist, «Systems on Chip» research laboratory, Omsk State Technical University (OmSTU, Omsk, Russia), viktor_erohin@mail.ru.

In this paper, the conductor model that can be used to design of various inductor layout configurations for any Si and SiGe technological processes is considered. The experimental prototypes of the test inductors were produced in the standard SiGe BiCMOS 130 nm process to verify the model. The chips measuring results showed that the characteristics of the prototypes taking into account the manufacturing tolerance are in the range of model simulated values. It has been found that the proposed model has a better convergence with the prototypes characteristics than 3D modeling. The equivalent circuit simulation speed can be orders of magnitude higher than the 3D simulation speed. The proposed model accuracy is achieved by taking into account the skin-effect and edge-effects in the dielectric and substrate. Using the skin-effect equivalent circuit the model can be run in Cadence Specter Simulator. It is necessary for the microwave LC-filters development.

Keywords: conductor model, Si, SiGe, inductor model, skin-effect, edge-effect.

References

- 1. Albert E. Ruehli, Giulio Antonini, Lijun Jiang. Skin-Effect Model for Round Wires in PEEC. International Symposium on Electromagnetic Compatibility - EMC EUROPE, Rome, Italy, 2012, 6 p.
- 2. Ren-Jia Chan, Jyh-Chyurn Guo. Analysis and Modeling of Skin and Proximity Effects for Millimeter-Wave Inductors Design in Nanoscale Si CMOS. *9th European Microwave Integrated Circuit Conference*, Rome, Italy, 2014, pp. 13-16.
- 3. F.M. Rotella, V. Blaschke, D. Howard. A Broad-Band Scalable Lumped-Element Inductor Model Using Analytic Expressions to Incorporate Skin Effect, Substrate Loss, and Proximity Effect. *Digest. International Electron Devices Meeting*, San Francisco, CA, USA, 2002, pp. 471-474.
- 4. Yu Cao, R.A. Groves, N.D. Zamdmer. Frequency-Independent Equivalent Circuit Model for On-chip Spiral Inductors. *Proceedings of the IEEE 2002 Custom Integrated Circuits Conference*, Orlando, FL, USA, 2002, pp. 217-220.
- 5. S. Kim, D.P. Neikirk. Compact Equivalent Circuit Model for the Skin Effect. *IEEE MTT-S International Microwave Symposium Digest*, San Francisco, CA, USA, 1996, pp. 1815-1818.
- 6. Mussa. Elsaadi, Mazhar B. Tayel, D. P. Steenson. An Empirical Formula of Fringing Field Capacitance for MEMS Tunable Capacitor Actuators. *IEEE 1st International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering MI-STA*, Tripoli, Libya, 2021, pp. 674-678.
- 7. Ehud Yariv. Edge corrections for parallel-plate capacitors. *European Journal of Applied Mathematics*, Cambridge, England, 2020, pp. 226-241.
- Ting Chen, Jin He, Pengwei Chen, Xuanyu Chen, Hao Wang, Sheng Chang, Qijun Huang. Accurate Modeling of Three-Port Center-tapped Octagonal Inductors for SPDT Switch Design in 0.13-μm BiC-MOS. *IEEE International Conference on Ubiquitous Wireless Broadband (ICUWB)*, Nanjing, China, 2016, 4 p.
- 9. Wei Yi Lim, M. Annamalai Arasu, M. Kumarasamy Raja. Modeling of Two Port Center-tapped to ground and Three Port scalable symmetrical inductor. *International Symposium on Integrated Circuits* (*ISIC*), Singapore, 2014, pp. 540-543.
- 10. Ji Chen, J.J. Liou. Improved and Physics-Based Model for Symmetrical Spiral Inductors. *IEEE Transactions on Electron Devices*, vol. 53, iss. 6, 2006, pp. 1300-1309.
- Amaya Goni, Javier del Pino, Benito Gonzalez, Antonio Hernandez. An Analytical Model of Electric Substrate Losses for Planar Spiral Inductors on Silicon. *IEEE Transactions on Electron Devices*, vol. 54, iss. 3, 2007, pp. 546-553.
- 12. J. Sathyasree, Venkata Vanukuru, Deleep R. Nair, Anjan Chakravorty. A Substrate Model for On-Chip Tapered Spiral Inductors with Forward and Reverse Excitations. *IEEE Transactions on Electron Devices*, vol. 66, iss. 1, 2019, 4 p.
- 13. Sathyasree Jeyaraman, Venkata Vanukuru, Deleep Nair, Anjan Chakravorty. Compact Modeling of Series Stacked Tapered Spiral Inductors. *IEEE 19th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems (SiRF)*, Orlando, FL, USA, 2019, 3 p.
- 14. Fabio Passos, M. Helena Fino, and Elisenda R. Moreno. Fully Analytical Characterization of the Series Inductance of Tapered Integrated Inductors. *International Journal of Electronics and Telecommunications*, Poland, 2014, pp. 73-77.
- 15. Maria Helena Fino. Using an Integrated Inductor Model in Ques. *Proceedings of the 21st International Conference Mixed Design of Integrated Circuits and Systems (MIXDES)*, Lublin, Poland, 2014, pp. 66-69.
- Monica Zolog, Dan Pitica, Ovidiu Pop. Characterization of Spiral Planar Inductors Built on Printed Circuit Boards. 30th International Spring Seminar on Electronics Technology (ISSE), Cluj-Napoca, Romania, 2007, pp. 308-313.
- 17. Y.K. Koutsoyannopoulos, Y. Papananos. Systematic Analysis and Modeling of Integrated Inductors and Transformers in RF IC Design. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 47, iss. 8, 2000, pp. 699-713.
- 18. Ji Chen, J.J. Liou. On-Chip Spiral Inductors for RF Applications: An Overview. *Journal of Semiconductor Technology and Science*, 2004, pp. 149-167.
- 19. Ahmed H. Shaltout, Stefano Gregori. Optimizing the Inductance Time-Constant Ratio of Polygonal Integrated Inductors. *IEEE 61st International Midwest Symposium on Circuits and Systems (MWSCAS)*, Windsor, ON, Canada, 2018, pp. 448-451.
- 20. Heng-Ming Hsu, Jen-Zien Chang, Hung-Chi Chien. Coupling Effect of On-Chip Inductor with Variable Metal Width. *IEEE Microwave and Wireless Components Letters*, vol. 17, iss. 7, 2007, pp. 498-500.