ISSN: 1813-324X

ТРУДЫ УЧЕБНЫХ ЗАВЕДЕНИЙ СВЯЗИ

Том 6. № 3
2020

Темы номера:

- Оценка эффективности космических аппаратов связи
- Выделение канального ресурса в гетерогенной сети радиодоступа
- Обработка пакетов данных в конвейерной системе

Vol. 6. Iss. 3
2020

PROCEEDINGS OF TELECOMMUNICATION UNIVERSITIES

Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Санкт-Петербургский государственный университет телекоммуникаций
им. проф. М.А. Бонч-Бруевича» (СПбГУТ)

Научный журнал

ТРУДЫ
УЧЕБНЫХ ЗАВЕДЕНИЙ СВЯЗИ

Том 6. № 3

Proceedings of Telecommunication Universities

Vol. 6. Iss. 3

Санкт-Петербург

2020

Труды учебных заведений связи. 2020. Т. 6. № 3. ISSN: 1813-324X (print)

Описание журнала

Научный журнал. Включен в Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук (распоряжение Минобрнауки РФ № 21-р от 12.02.2019), по специальностям:

05.11.07. Оптические и оптико-электронные приборы и комплексы
05.11.18. Приборы и методы преобразования изображений и звука
05.12.04. Радиотехника, в том числе системы и устройства телевидения
05.12.07. Антенны, СВЧ-устройства и их технологии
05.12.13. Системы, сети и устройства телекоммуникаций
05.12.14. Радиолокация и радионавигация
05.13.01. Системный анализ, управление и обработка информации
05.13.18. Математическое моделирование, численные методы и комплексы программ
05.13.19. Методы и системы защиты информации, информационная безопасность

Выпускается с 1960 года. Выходит 4 раза в год (ежеквартально). Издается на русском и английском языках.

Редакционный совет

Регистрационная информация

Журнал зарегистрирован Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций: ПИ № 77-77501 от 17.01.2020 г. (пред. рег. № 77-17986 от 07.04.2004 г.)

Подписной индекс по каталогу «Издания органов НТИ» Агентства «Роспечать»: 59983

Размещение в РИНЦ (elibrary.ru) по договору: № 59-02/2013R от 20.02.2013

Контактная информация

Учредитель и издатель: Федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича» (СПбГУТ)

Адрес редакции и издателя: 193232, Санкт-Петербург, пр. Большевиков, 22/1, к. 334/2
Тел.: +7 (812) 326-31-63, м. т. 2022, +79643759970
E-mail: tuzs@spbgut.ru
Web: http://tuzs.sut.ru
ВК: http://vk.com/spbtuzs

Proceedings of Telecommunication Universities. 2020. Vol. 6. Iss. 3. ISSN: 1813-324X (print)

Description

Scientific journal. The journal is included in the List of reviewed scientific publications, in which the main scientific results of dissertations for the degree of candidate of science and for the degree of doctor of science should be published (order of the Ministry of Education and Science of Russia No 21-r of 12 February 2019) in the field of:

05.11.07. Optical and optoelectronic devices and complexes
05.11.18. Devices and methods of transformation of images and sound
05.12.04. Radio engineering, including television systems and devices
05.12.07. Antennas, microwave devices and its technologies
05.12.13. Systems, networks and devices of telecommunications
05.12.14. Radiolocation and radio navigation
05.13.01. System analysis, management and information processing
05.13.18. Mathematical modelling, numerical methods and complexes of programs
05.13.19. Methods and systems of information security, cybersecurity

Since 1960. Published 4 times per year. Published in Russian and English.

Editorial Board

Registration Information

Registered by Federal Service for Supervision of Communications, Information Technology and Mass Media on 17/01/2020: PI No. 77-77501 (prev. reg. on 04/07/2004: No. 77-17986)

Subscription index for «NTI Editions» Agency «Rospechat» catalog: 59983

Accommodation in RINC (elibrary.ru) by agreement: № 59-02/2013R on 20.02.2013

Contact Information

Publisher: Federal State Budget-Financed Educational Institution of Higher Education «The Bonch-Bruevich Saint-Petersburg State University of Telecommunications» (SPbSUT)

Post address: 193232, Saint-Petersburg, Prospekt Bolshevikov, 22/1
Phone: +7 (812) 326-31-63, local 2022, +79643759970
E-mail: tuzs@spbgut.ru
Web: http://tuzs.sut.ru

СОДЕРЖАНИЕ CONTENTS

DOI:10.31854/1813-324X-2020-6-3

УДК 621.391 DOI:10.31854/1813-324X-2020-6-3-6-16

Методика оценивания эффективности применения космических аппаратов связи по целевому назначению

Е.С. Абазина[1], **А.А. Ковальский**[1*], **А.В. Питрин**[1]

[1]Военно-космическая академия имени А.Ф. Можайского,
Санкт-Петербург, 197198, Российская Федерация
[*]Адрес для переписки: vka@mil.ru

Информация о статье
Поступила в редакцию 29.06.2020
Принята к публикации 19.08.2020

Ссылка для цитирования: Абазина Е.С., Ковальский А.А., Питрин А.В. Методика оценивания эффективности применения космических аппаратов связи по целевому назначению // Труды учебных заведений связи. 2020. Т. 6. № 3. С. 6–16. DOI:10.31854/1813-324X-2020-6-3-6-16

Аннотация: *В статье представлен подход к сравнению космических аппаратов связи, который может быть применим при необходимости оперативного и обоснованного принятия решения о выборе космического аппарата связи в условиях ограничения времени на принятие решения. Заявленная методика выбора отличается от известных совокупным учетом характеристических показателей аппарата, существенно влияющих на качество выполнения им задач по предназначению. Оцениваемыми характеристиками в работе выбраны доступность, пропускная способность, помехозащищенность, надежность и управляемость аппарата. В статье приведены результаты апробации предлагаемой методики для определения наиболее эффективного аппарата связи в геостационарной орбитальной группировке, принадлежащей Российской Федерации, задействованной при организации информационного обмена подвижных абонентов на территории государства.*

Ключевые слова: *целевая эффективность, космический аппарат связи, методика оценивания, доступность, пропускная способность, помехоустойчивость, надежность, управляемость.*

1. Введение

В вопросах управления глобальными системами и информационного обмена с труднодоступными районами, при поддержке и сопровождении функционирования систем принятия решения и автоматизированных систем управления различного назначения, применение спутниковых систем связи (ССС) является приоритетным. Орбитальная группировка (ОГ) космических аппаратов связи (КАС) составляет основу ССС, ретрансляции и передачи данных. Несмотря на сравнительно молодую историю ССС, количество КАС в околоземном пространстве стремительно возрастает. Это справедливо как для геостационарных аппаратов, обеспечивающих круглосуточную видимость зоны обслуживания, так и для низкоорбитальных, ориентированных на предоставление спутниковой связи с объектами, находящимися в движении. Насыщенность низких и геостационарных орбит КАС разных поколений, обладающими различными энергетическими и надежностными характеристиками, определяет сложность задачи выбора единственного в качестве ретранслятора. Одним из подходов, позволяющим обосновать принятие решения такого выбора, является квалиметрический, который базируется на оценивании целевой эффективности каждого аппарата из множества оцениваемых.

2. Алгоритм оценивания целевой эффективности космических аппаратов связи

Согласно теории эффективности целенаправленных процессов, под эффективностью понимается комплексное операционное свойство («качество») целенаправленного процесса функционирования системы, характеризующее его приспособленность к достижению цели операции (к выполнению задачи системы) [1].

Поэтому под целевой (функциональной) эффективностью КАС в статье понимается степень его соответствия целевому назначению при применении (функционировании) в составе ОГ. Проведенные ис-

следования позволяют сделать вывод о том, что понятие «качества систем управления» является исходным по отношению к понятию «эффективность», которое в работах [1–3] полагается следствием «качества» для случаев, когда система обладает целесообразностью. При этом качество имеет значение некоторой совокупности отдельных полезных свойств, разделяемых по иерархическим уровням в зависимости от степени их общности. Предметом изучения квалиметрии являются количественные оценки качества объекта или процесса, которые характеризуют степень их соответствия предъявляемым требованиям (или целям). Научно-методический аппарат квалиметрии, послужил базой для разработки представляемой методики оценивания целевой эффективности КАС. Предлагаемая методика может быть применена для получения обоснованной оперативной, но приблизительной оценки некоторого КАС по отношению ко множеству КАС, участвующих в оценивании между собой.

Схематично предлагаемую методику оценивания целевой эффективности КАС можно представить в виде алгоритма на рисунке 1.

Реализация предложенного алгоритма оценивания целевой эффективности КАС предполагает проведение ряда операций (шагов), на каждом из которых производится оценивание одной из характеристик аппарата, влияющих на возможность выполнения аппарата задач по предназначению с заданным качеством. Расчет характеристических показателей может быть реализован как последовательно, так и параллельно, что определяется количеством областей буферной памяти, их размерами, а также быстродействием применяемых процессорных устройств.

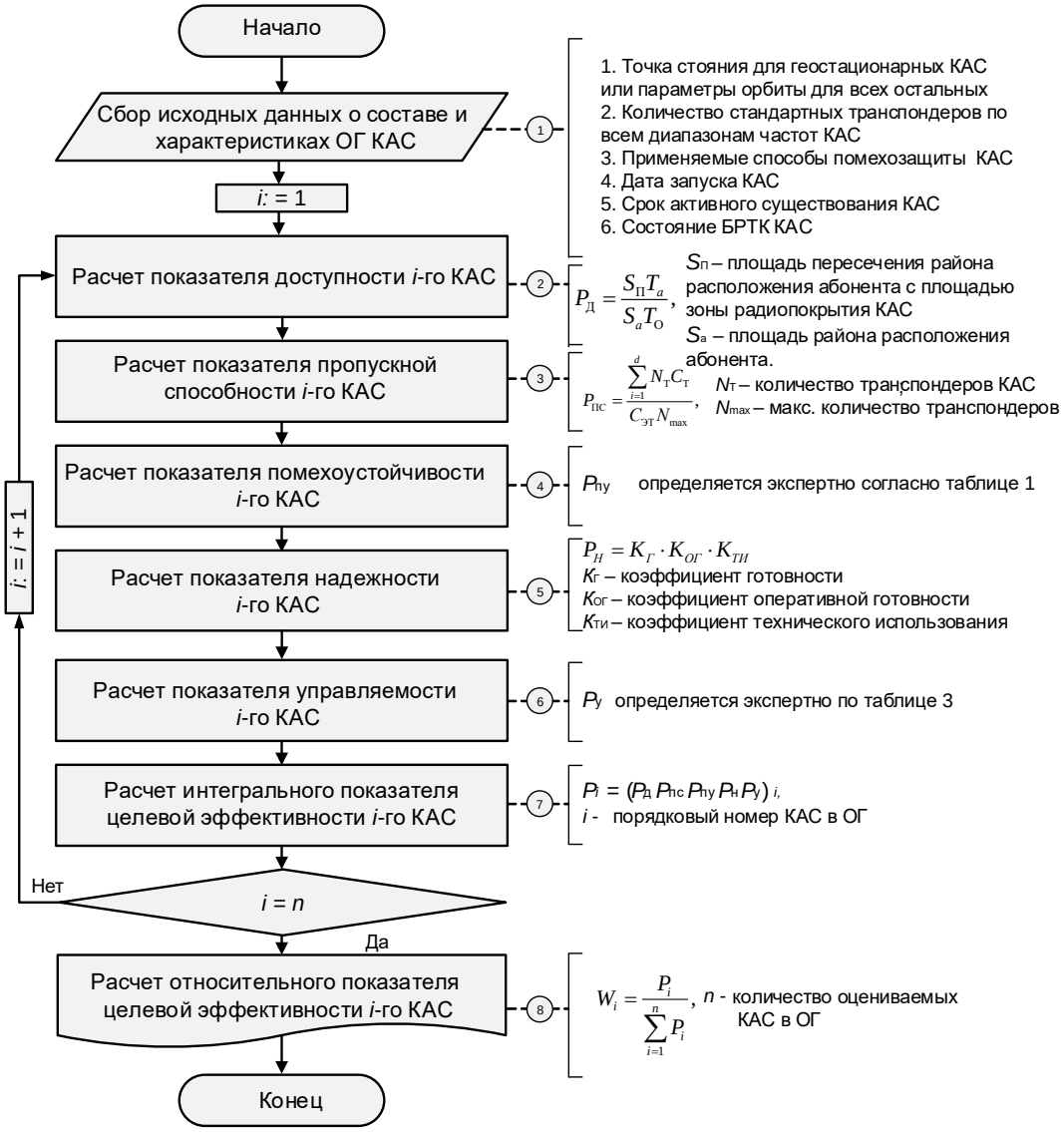


Рис 1. Блок-схема алгоритма оценивания целевой эффективности КАС

Fig. 1. Algorithm's Flowchart Evaluating the Communication Satellite's Target Efficiency

3. Состав основных показателей, исходных и выходных данных при расчете целевой эффективности космических аппаратов связи

Для оценивания целевой эффективности КАС необходимо определить состав исходных и выходных данных, а также показателей и критериев, по которым будет производиться их сравнение между собой.

В качестве исходных данных, собираемых на шаге 1, предлагается использовать следующие технические характеристики КАС:

– точка стояния для геостационарных или параметры орбиты для всех остальных КАС;

– количество стандартных транспондеров по всем диапазонам частот;

– применяемые способы повышения помехоустойчивости;

– дата начала эксплуатации;

– срок активного существования;

– состояние бортового ретрансляционного комплекса (БРТК);

– применяемые способы повышения надежности.

В качестве выходных данных методики и основного оцениваемого показателя, по которому выносится решение, в работе выбран показатель относительной целевой эффективности КАС (W_i), который показывает целевую эффективность отдельного аппарата относительно всех рассматриваемых в составе ОГ. В общем случае оцениваемый КАС может и не принадлежать к составу ОГ. В соответствии с [1–4] W_i принимает значения в диапазоне $W_i \in [0; 1]$ и может быть рассчитан по формуле (1) на шаге 8 приведенного алгоритма:

$$W_i = \frac{P_i}{\sum_{i=}^{n} P_i}, \qquad (1)$$

где W_i – относительный показатель целевой эффективности i-го КА из множества оцениваемых; i – условный порядковый номер КА во множестве оцениваемых, $i = \overline{1,n}$; P_i – значение интегрированного показателя целевой эффективности i-го КАС, $P_i \in [0; 1]$.

Основными характеристиками КАС с позиций выполнения задач по целевому предназначению являются:

– доступность;

– пропускная способность;

– помехозащищенность;

– надежность;

– управляемость.

Предлагаемая методика предполагает по мере сбора исходных данных вычисление соответствующих характеристических показателей (P_i) с их последующей нормировкой.

Рассмотрим наиболее значимые характеристики аппарата, оцениваемые в методике, определим показатели и критерии их эффективности, а также способы расчета.

4. Показатель доступности космических аппаратов связи и критерии оценивания эффективности

Доступность КАС – способность аппарата обеспечивать абонентам спутниковой сети связи доступ к радиоресурсу при сохранении назначенных приоритетов и способов установления связи [5]. Одним из наиболее часто вычисляемых показателей доступности КАС является вероятность санкционированного доступа абонента независимо от его местоположения относительно радиоресурса. Очевидно, что вероятность санкционированного доступа зависит от взаимного расположения абонента спутниковой сети связи и зоны радиопокрытия КАС, что фактически определяется вероятностью нахождения абонента в области обслуживания аппарата и зависит от выполнения условия взаимной радиовидимости спутникового абонентского терминала и КАС (рисунок 2).

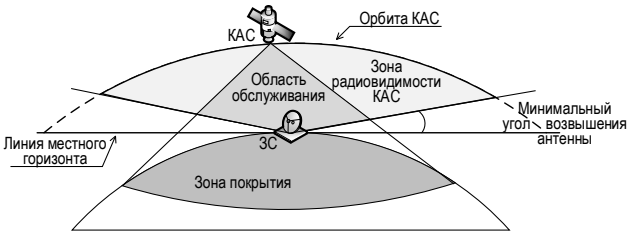


Рис. 2. Взаимосвязь зон покрытия, обслуживания и радиовидимости КАС

Fig. 2. The Relationship of Coverage Zone, Service Zone and Radio Visibility of Communication Satellite

Приняв допущение о том, что все абоненты равномерно распределены по зоне расположения, проведем расчет показателя доступности КАС ($P_Д$). При наличии подвижных абонентов, для этого необходимо вычислить коэффициент пересечения зоны радиопокрытия и района расположения абонента, а также времени нахождения аппарата и абонентов в зоне взаимной радиовидимости.

Для расчета показателя доступности КАС, выполняемого на шаге 2, используем формулу:

$$P_Д = \frac{S_П T_a}{S_a T_O}, \qquad (2)$$

где $P_Д$ – показатель доступности КАС; $S_П$ – площадь пересечения района расположения абонента с площадью зоны радиопокрытия КАС; S_a – площадь района расположения абонента; T_a – период времени, в течение цикла обращения КАС вокруг Земли, когда абонент находится в зоне радиопокрытия КАС; T_O – период времени обращения КАС вокруг Земли.

Очевидно, что частным случаем при размещении КАС на геостационарной орбите отношение будет равно 1, т. к. абоненты всегда находятся в одном и том же положении относительно аппарата. В случае, когда районов расположения абонентов несколько, расчет ведется по пересечению всех рассматриваемых районов.

Для случая, когда рассматриваются только стационарные абоненты, расчет показателя доступности КАС $P_Д$ ведется по упрощенной формуле и определяется отношением количества обслуживаемых аппаратом абонентов, находящихся в зоне радиопокрытия, к их общему количеству.

5. Показатель пропускной способности космических аппаратов связи и критерии оценивания эффективности

Пропускная способность КАС – способность аппарата, определяющая совокупный объем информации, передаваемый и обрабатываемый в образуемых им радиоканалах за единицу времени, и может быть количественно оценена матрицей пропускных способностей направлений связи. В этом случае ее принято рассчитывать через суммарное количество эквивалентных (стандартных) транспондеров в заданных диапазонах частот [6, 7]. В качестве величины эквивалентного транспондера КАС принята полоса частот, равная 36 МГц. В этом случае нормированный показатель пропускной способности КАС ($P_{ПС}$) вычисляется по формуле (3) на шаге 3 (при последовательном исполнении алгоритма методики):

$$P_{ПС} = \frac{\sum_{i=1}^{d} N_т C_т}{C_{ЭТ} N_{max}}, \quad (3)$$

где d – количество диапазонов частот, в которых может функционировать КАС; $N_т$ – количество транспондеров КАС в заданном диапазоне частот; $C_т$ – полоса частот транспондеров КАС в заданном диапазоне; $C_{ЭТ}$ – полоса частот эквивалентного транспондера КАС; N_{max} – максимальное число транспондеров КАС в рассматриваемом множестве КАС (или в ОГ).

6. Показатель помехоустойчивости космических аппаратов связи и критерии оценивания эффективности

Помехоустойчивость КАС – способность аппарата выполнять задачи по предназначению в условиях воздействия всех видов помех [5].

Выполнить грубое (приблизительное, неточное) оценивание помехоустойчивости КАС возможно в результате сравнения современных стандартов спутниковой связи, реализуемых в КАС. Помехоустойчивость систем связи находится в жестком соотношении со скоростью передачи информации и напрямую связана с вопросами оптимального использования спектрального ресурса спутникового канала. Современные стандарты спутниковой связи предусматривают возможности адаптивного изменения баланса между помехоустойчивостью и скоростью передачи информации за счет выбора модуляции и кодирования. Основными стандартами, реализуемыми в действующих КАС на современном этапе, являются DVB-RCS (*от англ.* Digital Video Broadcasting – Return Channel via Satellite), DVB-S/S2 (*от англ.* Digital Video Broadcasting – Sputnik), а также режим псевдослучайной перестройки рабочей частоты (ППРЧ) и использование фазоманипулированных шумоподобных сигналов (ФМ ШПС), сравнение которых выполнено в работе [8]. Выводы авторов о том, что стандарт DVB-S является менее помехоустойчивым, чем DVB-S2 или DVB-RCS. Применение ППРЧ и ФМ ШПС позволяет достичь наилучшей помехоустойчивости, что обосновывает их использование в методике для расчета показателя помехоустойчивости КАС по методу экспертных оценок. Также при расчете учтена возможность применения и других способов обеспечения помехоустойчивости: обработка сигнала на борту, использование многолучевых антенн, управление лучами антенны, формирование «нулей» диаграммы направленности антенны.

Вычисление показателя помехоустойчивости КАС производится на шаге 4 (при последовательном исполнении алгоритма методики) по формуле (4), позволяющей определить соотношение наличия и эффекта технологий обеспечения помехоустойчивости БРТК КАС, и искомого показателя:

$$P_{ПУ} = \sum_{i=1}^{m} K_i, \quad (4)$$

где m – количество технологий, влияющих на помехоустойчивость БРТК КАС; K_i – показатель эффекта технологий обеспечения помехоустойчивости БРТК КАС, определяемый экспертно по таблице 1.

ТАБЛИЦА 1. Технологии обеспечения помехоустойчивости КАС

TABLE 1. Communication Satellite's Noise Immunity Technologies

№ п/п	Технологии обеспечения помехоустойчивости БРТК КАС	K_i
1.	Стандарт DVB-S	0,07/0,09
	Стандарты DVB-S2/ DVB-RCS	
2.	Использование ФМ ШПС	0,18
3.	Режим ППРЧ	0,22
4.	Многолучевые антенны	0,10
5.	Режим обработки сигналов на борту	0,14
6.	Управление лучами антенны	0,11
7.	Формирование «нулей» диаграммы направленности антенны	0,16
8.	Отсутствие технологий	0,05

7. Показатель надежности космических аппаратов связи и критерии оценивания эффективности

Надежность КАС – способность КАС обеспечивать спутниковую связь, сохраняя во времени значение эксплуатационных показателей в пределах норм эксплуатации.

Поэтому надежность КАС определяется совокупностью свойств, характеризующих способность КАС

обеспечивать в процессе функционирования получение заданного в тактико-техническом задании выходного эффекта при заданных условиях и режимах эксплуатации [8–11].

Для показателей надежности космических систем и комплексов в соответствии с [8–11] введена классификация, выделяющая обобщенные, основные и дополнительные показатели. Для оценивания надежности при заданных способах и условиях применения вычисляют обобщенные показатели, которые характеризуют надежность на всех периодах эксплуатации. Основные показатели надежности необходимо получать для оценивания в отдельные периоды эксплуатации. При решении частных задач обеспечения, оценки и контроля надежности используют дополнительные показатели.

В методике для оценивания надежности КАС используются обобщенные комплексные показатели [9] – коэффициенты готовности $K_{\text{Г}}$, оперативной готовности $K_{\text{ОГ}}$ и технического использования $K_{\text{ТИ}}$, на основании которых рассчитывается итоговый показатель надежности КАС $P_{\text{Н}}$ (шаг 5 последовательного исполнения алгоритма методики):

$$P_{\text{Н}} = K_{\text{Г}} \cdot K_{\text{ОГ}} \cdot K_{\text{ТИ}}. \qquad (5)$$

Коэффициент готовности $K_{\text{Г}}$ характеризует готовность КАС к применению по назначению только в произвольный момент времени.

Оценивание готовности КАС в соответствии с [10] представляет собой формализованную процедуру анализа конструкции изделия, включающую в себя оценку принятых конструктивных решений на предмет соответствия их установленным требованиям к готовности или восстанавливаемости. В методике расчет коэффициента внутренней готовности осуществляется по формуле (6), каждое из слагаемых определяется экспертно (таблица 2):

$$K_{\text{Г}} = \sum_{g=1}^{G} K_{\text{Г}_g}, \qquad (6)$$

где $K_{\text{Г}_g}$ – экспертное значение коэффициента готовности при реализации g-го решения обеспечения восстанавливаемости КАС; G – число реализуемых решений обеспечения восстанавливаемости КАС.

Коэффициент оперативной готовности $K_{\text{ОГ}}$ характеризует надежность КАС, необходимость применения которого возникает в произвольный момент времени. Коэффициент оперативной готовности имеет вероятностный характер и рассчитывается как вероятность того, что КАС окажется в работоспособном состоянии в данный момент времени и, начиная с этого момента, будет работать безотказно в течение заданного интервала времени или, другими словами, это вероятность безотказной работы КАС на заданный период времени [8–11].

Коэффициент оперативной готовности $K_{\text{ОГ}}^{\text{стат}}$ может быть статистически рассчитан по формуле [8–11]:

$$K_{\text{ОГ}}^{\text{стат}} = \frac{N(t_\infty + \Delta t)}{N(0)}, \qquad (7)$$

где $N(t_\infty + \Delta t)$ – число критически важных блоков КАС, работоспособных в произвольный, достаточно удаленный момент времени t_∞, и проработавших безотказно в период времени Δt; $N(0)$ – общее число критически важных блоков КАС.

Под критически важными блоками КАС понимаются блоки, отказ, которых приводит к отказу БРТК КАС и невозможности его функционирования по целевому назначению.

ТАБЛИЦА 2. Решения, обеспечивающие восстанавливаемость КАС

TABLE 2. Solutions that Ensure Communications Satellites's Recoverability

g	Возможность установки переключателей резерва	$K_{\text{Г}_g}$
1	*Структурное резервирование* – способ обеспечения надежности изделия за счет введения в состав конструкции изделия дополнительных (резервных) составных частей, способных выполнять функции основных составных частей при их отказе	0,30
2	*Функциональное резервирование* – способ обеспечения надежности изделия за счет применения многофункциональности отдельных составных частей изделия, способных выполнять функции других отказавших составных частей, без существенного ухудшения функциональных характеристик изделия; определяет возможность динамического перераспределения функциональной нагрузки элементов при изменении функциональной структуры изделия	0,25
3	*Алгоритмическое резервирование* – способ обеспечения надежности изделия за счет введения в состав программного обеспечения дополнительных, параллельно работающих ветвей алгоритмов минимально возможной сложности; определяет возможность включения в состав систем управления алгоритмов и средств, осуществляющих перестройку структуры (реконфигурацию), а также способных использовать для выполнения задания работоспособные ресурсы	0,20
4	*Временное резервирование* – способ обеспечения надежности изделия за счет создания временных перерывов в работе отдельных составных частей изделия, которые при их отказе могут быть использованы для восстановления, без нарушений выполнения целевых задач изделием	0,15
5	*Информационное резервирование* – способ обеспечения надежности изделия путем создания избыточных копий массивов информации или семантически адекватных источников информации, за счет введение дополнительной информации, предназначенной для восстановления основной информации в случае ее искажения; способ определяет возможность распараллеливания работ, связанных с функционированием изделия	0,10
6	*Отсутствие резервирования*	0,05

Чаще всего для КАС коэффициент оперативной готовности $K_{ОГ}$ находится в рамках 0,70–0,95 на период гарантийного срока активного существования.

Коэффициент технического использования $K_{ТИ}$ характеризует долю времени нахождения КАС в работоспособном состоянии относительно общей продолжительности эксплуатации. В предлагаемой методике $K_{ТИ}$ определяется через коэффициент остаточного ресурса КАС ($K_{ОР}$), который в обобщенном виде равен отношению периода нахождения КАС в эксплуатации к проектируемому гарантийному сроку активного существования и рассчитывается по формуле:

$$K_{ОР} = \frac{T_Э}{T_{гар}}, \quad (8)$$

где $T_Э$ – период нахождения КАС в эксплуатации; $T_{гар}$ – проектируемый гарантийный срок активного существования КАС.

Далее коэффициент $K_{ТИ}$ КАС определяется экспертно в соответствии с таблицей 3.

ТАБЛИЦА 3. Соотношение коэффициента технического использования с его остаточным ресурсом

TABLE 3. Correlation of the Communications Satellite's Technical Use Coefficient with its Remaining Resource

№ п/п	$K_{ОР}$	$K_{ТИ}$
1.	0–0,2	1,00
2.	0,2–0,4	0,95
3.	0,4–0,6	0,90
4.	0,6–0,8	0,85
5.	0,8–1,0	0,80
6.	0,1–1,2	0,75
7.	1,2–1,4	0,70
8.	1,4–1,6	0,65
9.	1,6–1,8	0,60
10.	1,8–2,0	0,55
11.	2,0–∞	0,50

8. Показатель управляемости космических аппаратов связи и критерии оценивания эффективности

Управляемость КАС – это способность аппарата изменять свое состояние в заданных пределах при воздействиях на нее органов управления в соответствии с изменениями обстановки [5]. Оценивание управляемости КАС выполняют по значению вероятности того, что время выполнения задач по управлению им не превысит допустимой длительности технологического цикла управления. Сложность расчета данного показателя в режиме масштаба времени, приближенного к реальному, определяет выбор экспертного способа оценивания в предлагаемой методике.

В соответствии с [11–13] на взаимодействие бортовой системы управления аппарата, наземный автоматизированный комплекс управления и целевой бортовой аппаратуры связи немалое влияние оказывает тип орбиты. Увеличение расстояния от наземного автоматизированного комплекса управления до КАС влечет ухудшение качества радиосигнала в прямом и обратном направлении (на линии «вверх» и на линии «вниз»), что ведет к задержкам в управлении, уменьшению скорости передачи информации, следовательно, и к снижению объемов телеметрической информации от целевой бортовой аппаратуры.

В связи с этим оперативное приближенное оценивание управляемости КАС $P_У$ определяется его типом орбиты на шаге 6 алгоритма представляемой методики согласно таблице 4.

ТАБЛИЦА 4. Взаимосвязь показателя управляемости КАС и типа его орбиты

TABLE 4. The Relationship between the Communication Satellite's Controllability Index and the Type of its Orbit

№ п/п	Тип орбиты КАС	$P_У$
1.	Низкая круговая	1,00
2.	Средняя круговая	0,95
3.	Геостационарная	0,90
4.	Высокоэллиптическая	0,85

Экспертные оценки параметров, представленные в таблицах 1–4, сформированы на основе анализа материалов работ [8–13].

9. Расчет показателя целевой эффективности космических аппаратов связи

Для обобщения полученных характеристических показателей КАС на шаге 7 выполняется расчет интегрального показателя целевой эффективности КАС по формуле:

$$P_i = (P_Д \cdot P_{ПС} \cdot P_{ПУ} \cdot P_Н \cdot P_У)_i, \quad (9)$$

где i – условный порядковый номер КАС во множестве оцениваемых (или в составе рассматриваемой ОГ).

На практике также удобно использовать нормированный показатель целевой эффективности КАС $P_i^{норм}$, который рассчитывается по формуле:

$$P_i^{норм} = P_i / P_{max}, \quad (9)$$

где $P_i^{норм}$ – нормированное значение интегрированного показателя целевой эффективности i-го КАС, $P_i \in [0; 1]$; P_{max} – максимальное значение интегрированного показателя целевой эффективности КАС, входящего в состав ОГ, относительно которого будет производится сравнение других аппаратов.

Нормированный показатель эффективности отображает, на сколько применение оцениваемого КАС соответствует целевому назначению относительно эталонного (самого наилучшего) аппарата, входящего в состав рассматриваемой ОГ.

На заключительном этапе работы представленного алгоритма производится расчет относительного показателя целевой эффективности КАС (шаг 8), который характеризует эффективность применения отдельного аппарат относительно всей орбитальной группировки при организации информационного обмена. Расчет осуществляется в соответствии с формулой (1) и завершает алгоритм методики оценивания целевой эффективности КАС.

10. Апробация методики оценивания эффективности применения космических аппаратов связи по целевому назначению

Апробация предложенной методики была проведена для орбитальной группировки, состоящей из 24 КАС. Для оценивания целевой эффективности выбраны геостационарные аппараты, принадлежащие РФ, серий «Ямал», «Экспресс», «Радуга» и «Луч». Цель апробации – оценить эффективность применения КАС при организации информационного обмена подвижных абонентов на территории РФ и определить наиболее эффективные из них.

Исходные данные для проведения расчетов согласно разработанной методике представлены в таблице 5. После обобщения исходных данных, произведем расчеты показателей целевой эффективности КАС по предложенному алгоритму расчета (рисунок 1).

Сводные данные характеристических показателей КАС и результаты расчета их целевой эффективности приведены в таблице 6. В результате проведенных расчетов построим диаграммы распределения $P_i^{норм}$ и относительного показателя целевой эффективности КАС W_i, изображенных на рисунках 3 и 4, соответственно.

ТАБЛИЦА 5. Исходные данные оцениваемых КАС для апробации предложенной методики

TABLE 5. Initial Data of the Estimated Communication Satellites for Testing the Proposed Method

№ п/п	Наименование КАС	Технические характеристики оцениваемых КАС			
		Точка стояния, град.	Количество транспондеров	Дата запуска	Срок активного существования, лет
1.	Экспресс-АМУ1	36	70	24.12.2015	15
2.	Экспресс-А1Р	145	10	10.06.2002	10
3.	Экспресс-АМ2	145	29	29.03.2005	12
4.	Экспресс-АМ22	80	30	28.12.2003	12
5.	Экспресс-АМ3	103	29	24.06.2005	12
6.	Экспресс-АМ33	96,5	27	28.01.2008	12
7.	Экспресс-АМ44	-11	27	11.02.2009	12
8.	Экспресс-АМ5	140	84	26.12.2013	15
9.	Экспресс-АМ6	53	72	21.10.2014	15
10.	Экспресс-АМ7	40	62	18.03.2015	15
11.	Экспресс-АМ8	-14	42	14.09.2015	15
12.	Экспресс-АТ1	56	32	15.03.2014	15
13.	Экспресс-АТ2	140	16	15.03.2014	15
14.	Луч-5А	167	21	11.12.2011	10
15.	Луч-5Б	-16	21	02.11.2012	10
16.	Луч-5В	95	21	28.04.2014	10
17.	Луч-5	95	21	27.09.2014	15
18.	Радуга-1М1	89,8	24	28.02.2009	10
19.	Радуга-1М2	70	24	28.01.2010	10
20.	Радуга-1М3	85	24	11.11.2013	10
21.	Ямал-202	49	18	24.11.2003	12
22.	Ямал-300К	90	26	02.11.2012	11
23.	Ямал-401	90	53	15.12.2014	15
24.	Ямал-402	55	46	08.12.2012	11

ТАБЛИЦА 6. Выходные данные оцениваемых КАС для апробации предложенной методики

TABLE 6. Output Data of Estimated Communication Satellites for Testing the Proposed Method for Testing the Proposed Method

№ п/п	Наименование	Характеристические показатели КАС					$P_i^{норм}$, %	W_i, %
		$P_д$	$P_{пс}$	$P_{пу}$	$P_н$	$P_у$		
1.	Экспресс-АМУ1	0,48	0,83	0,19	0,92	0,90	78	7,5
2.	Экспресс-А1Р	0,47	0,12	0,07	0,57	0,90	3	0,2
3.	Экспресс-АМ2	0,47	0,35	0,07	0,73	0,90	9	0,9
4.	Экспресс-АМ22	0,63	0,35	0,07	0,71	0,90	12	1,2
5.	Экспресс-АМ3	0,63	0,32	0,17	0,72	0,90	28	2,7
6.	Экспресс-АМ33	0,18	0,32	0,17	0,74	0,90	8	0,8
7.	Экспресс-АМ44	0,50	1,00	0,19	0,76	0,90	82	7,9
8.	Экспресс-АМ5	0,58	0,86	0,19	0,88	0,90	94	9,1
9.	Экспресс-АМ6	0,50	0,74	0,19	0,89	0,90	71	6,8
10.	Экспресс-АМ7	0,16	0,50	0,19	0,91	0,90	16	1,5
11.	Экспресс-АМ8	0,60	0,38	0,19	0,92	0,90	45	4,4
12.	Экспресс-АТ1	0,50	0,19	0,19	0,90	0,90	18	1,8
13.	Экспресс-АТ2	0,33	0,25	0,19	0,90	0,90	16	1,5
14.	Луч-5А	0,15	0,25	0,21	0,81	0,90	7	0,7
15.	Луч-5Б	0,63	0,25	0,21	0,82	0,90	30	2,9
16.	Луч-5В	0,63	0,25	0,21	0,86	0,90	32	3,1
17.	Луч-5	0,63	0,29	0,21	0,83	0,90	35	3,4
18.	Радуга-1М1	0,63	0,29	0,65	0,76	0,90	100	9,6
19.	Радуга-1М2	0,63	0,29	0,65	0,74	0,90	97	9,4
20.	Радуга-1М3	0,56	0,21	0,65	0,85	0,90	75	7,2
21.	Ямал-202	0,63	0,31	0,07	0,70	0,90	11	1,0
22.	Ямал-300К	0,63	0,63	0,19	0,82	0,90	70%	6,7
23.	Ямал-401	0,59	0,55	0,19	0,90	0,90	63	6,1
24.	Ямал-402	0,63	0,36	0,19	0,81	0,90	39	3,7

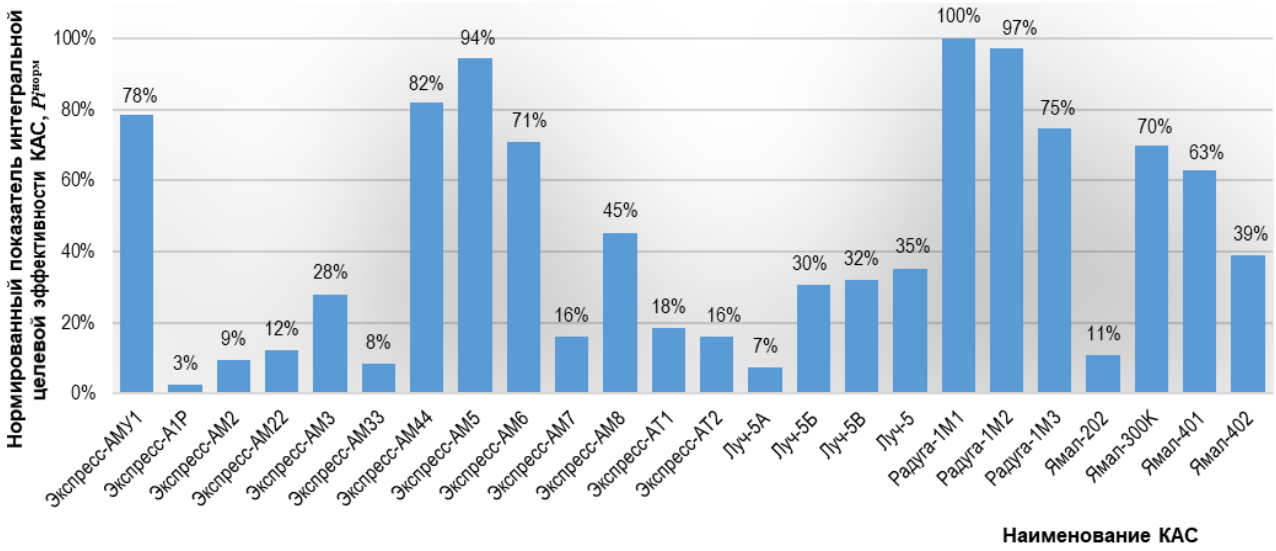


Рис 3. Диаграмма распределения нормированного показателя целевой эффективности КАС в составе рассматриваемой ОГ

Fig. 3. Distribution Diagram of the Communication Satellite's Target Efficiency Normalized Indicator as a Part of the Considered Orbital Grouping

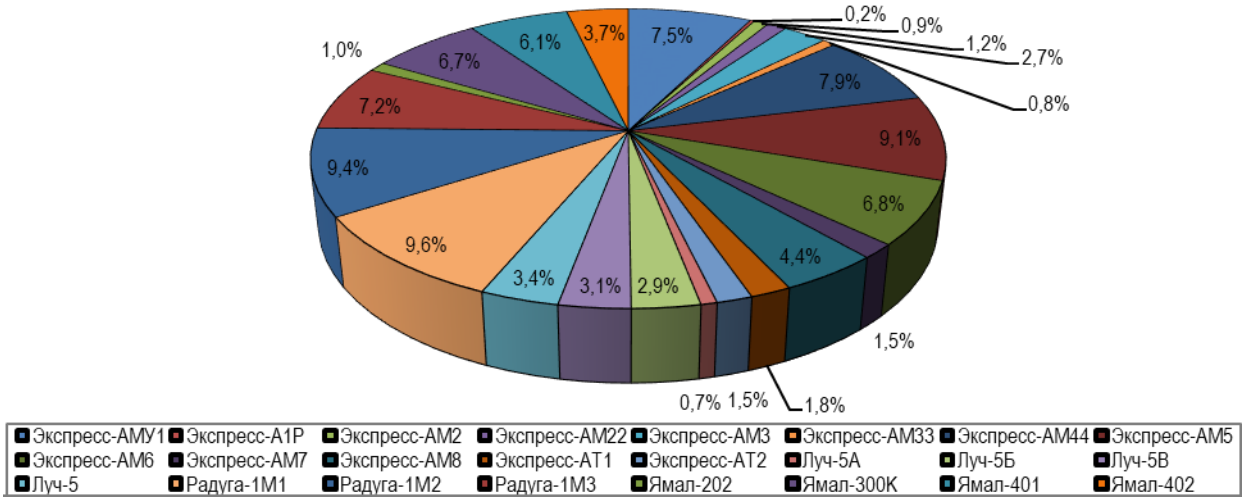


Рис 4. Диаграмма распределения относительного показателя целевой эффективности КАС в составе рассматриваемой ОГ

Fig. 4. Distribution Diagram of the Communication Satellite's Target Efficiency Relative Indicator as a Part of the Considered Orbital Grouping

Анализ приведенных результатов показывает, что при организации информационного обмена подвижных абонентов на территории РФ наиболее эффективно применение следующих КАС:

за счет высоких показателей доступности и помехоустойчивости

Радуга-1М1 (показатели $P_i^{норм}$ = 100 %, W_i = 9,6 %);

Радуга-1М2 (показатели $P_i^{норм}$ = 97 %, W_i = 9,4 %);

за счет высоких показателей доступности и пропускной способности

Экспресс-АМ5 (показатели $P_i^{норм}$ = 94 %, W_i = 9,1 %);

Экспресс-АМ4 (показатели $P_i^{норм}$ = 82 %, W_i = 7,9 %).

КАС серии «Ямал» показали среднюю эффективность в составе ОГ:

Ямал-300К (показатели $P_i^{норм}$ = 70 %, W_i = 6,7 %);

Ямал-401 (показатели $P_i^{норм}$ = 63 %, W_i = 6,1 %).

КАС серии «Луч» показали низкую эффективность в составе ОГ, что объясняется тем, что основной задачей данных КАС является ретрансляция инфомации от других космических аппаратов:

Луч-5 (показатели $P_i^{норм}$ = 35 %, W_i = 3,4 %);

Луч-5В (показатели $P_i^{норм}$ = 32 %, W_i = 3,1 %).

10. Заключение

Представленная в работе методика оценивания целевой эффективности базируется на положениях квалиметрии и ориентирована на расчет интегрального и относительного показателей эффективности, учитывающего характеристические показатели КАС. В качестве основных характеристик, подлежащих оцениванию, в методике выбраны доступность, пропускная способность, помехоустойчивость, надежность и управляемость КАС.

Данная методика может быть использована должностными лицами при:

– планировании и распределении орбитального канального ресурса сетей спутниковой связи;

– резервировании в целях обеспечения устойчивости направлений спутниковой связи;

– сравнении космических аппаратов связи, когда требуется оперативное и обоснованное принятие решения о выборе одного из них для выполнения целевых задач по обеспечению связью широкого круга абонентов;

– обосновании тактико-технических характеристик космических аппаратов связи.

Методика является выражением ситуационного подхода к принятию решений в условиях жестких временных ограничений на всесторонний анализ оцениваемого объекта или недостаточности исходных данных и может быть рекомендована для применения в интересах обеспечения требуемого уровня оперативности на выработку решения. При этом получаемые оценки являются приближенными, а при возможности учета дополнительных факторов, условий и увеличения интервала оценивания могут быть уточнены и скорректированы.

Наземный сегмент космической системы в статье не рассматривается, что составляет ограничение работы. Это обоснованно ориентацией методики на сравнение космических аппаратов между собой при необходимости оперативного выбора одного из них.

В данной статье, посвященной лишь оцениванию эффективности применения космических аппаратов связи по целевому назначению, не ставилось задачи оценивания эффективности системы спутниковой связи в целом. Однако предложенная методика может быть применена для ее решения как один из компонентов более сложного научно-методического аппарата.

Изложенный в статье научно-методический подход в совокупности с проведенным системным анализом может быть использован для оценивания эффективности систем спутниковой связи между собой, что позволит более обоснованно определить оптимальный состав и тактико-технических характеристики разрабатываемых систем. [14–16].

Список используемых источников

1. Петухов Г.Б. Основы теории эффективности целенаправленных процессов. Часть 1. Методология, методы, модели. М.: МО СССР, 1989. 660 с.
2. Фомин Л.А., Будко П.А. Эффективность и качество инфокоммуникационных систем. Методы оптимизации: Монография. М.: ООО ИФ «Физико-математическая литература», 2008. 296 с.
3. Солодкая М.С. Надежность, эффективность, качество систем управления // QUALITY.EUP.RU. URL: https://www.quality.eup.ru/MATERIALY10/qsm.htm (дата обращения 13.08.2020)
4. Мануйлов Ю.С., Павлов А.Н., Новиков Е.А. Системный анализ и организация автоматизированного управления космическими аппаратами. СПб.: ВКА имени А.Ф. Можайского, 2010. 266 с.
5. ГОСТ Р 53801-2010 Связь федеральная. Термины и определения. Национальный стандарт Российской Федерации. М.: Стандартинформ, 2011.
6. Жуков С.Е., Ковальский А.А., Квасов М.Н., Митряев Г.А. Оперативное распределение радиоресурса системы спутниковой связи в целях обеспечения управления космическими аппаратами // Труды Научно-исследовательского института радио. 2017. № 2. С. 29–36.
7. Зиннуров С.Х., Ковальский А.А., Митряев Г.А. Решение задачи оптимального планирования для сеансов управления орбитальной группировкой космических аппаратов // Труды учебных заведений связи. 2018. Т. 4. № 1. С. 67–74. DOI:10.31854/1813-324x-2018-1-67-74
8. Паршуткин А.В., Маслаков П.А. Исследование помехоустойчивости современных стандартов спутниковой связи к воздействию нестационарных помех // Труды СПИИРАН. 2017. № 4(53). С. 159–177. DOI:10.15622/sp.53.8
9. ГОСТ Р 56526-2015 Требования надежности и безопасности космических систем, комплексов и автоматических космических аппаратов единичного (мелкосерийного) изготовления с длительными сроками активного существования. М.: Стандартинформ, 2016.
10. ГОСТ Р 53111-2008 Устойчивость функционирования сети связи общего пользования. Требования и методы проверки. М.: Стандартинформ, 2019.
11. ГОСТ Р 58625-2019 Системы и комплексы космические. Анализ ремонтопригодности. Общие требования. М.: Стандартинформ, 2020.
12. ГОСТ Р 58628-2019 Системы и комплексы космические. Анализ готовности. Общие требования. М.: Стандартинформ, 2020.
13. Бортовые системы управления космическими аппаратами / Под ред. А.С. Сырова. М.: Изд-во МАИ-ПРИНТ, 2010. 304 с.
14. Новиков Е.А. Гибкие технологические стратегии управления целевым применением космической навигационной системы. СПб: ВКА им. А.Ф. Можайского, 2009. 175 с.
15. Мануйлов Ю.С., Птушкин А.И., Стародубов В.А. Методологические основы применения гибких стратегий управления космическими аппаратами. СПб: МО РФ, 2002. 102 с.
16. Мануйлов Ю.С., Калинин В.Н., Гончаревский В.С. и др. Управление космическими аппаратами и средствами наземного комплекса управления: учебник. СПб: ВКА им. А.Ф. Можайского, 2010. 609 с.

* * *

The Communication Satellites's Target Effectiveness Evaluating Methodology

E. Abazina[1], **A. Kovalsky**[1], **A. Pitrin**[1]

[1]Military space academy of A.F. Mozhaysky,
St. Petersburg, 197198, Russian Federation

Article info
DOI:10.31854/1813-324X-2020-6-3-6-16
Received 29th June 2020
Accepted 19th August 2020

For citation: Abazina E., Kovalsky A., Pitrin A. The Communication Satellites's Target Effectiveness Evaluating Methodology. *Proc. of Telecom. Universities.* 2020;6(3):6–16. (in Russ.) DOI:10.31854/1813-324X-2020-6-3-6-16

Abstract: *There is an approach of the communication satellites comparison presents in the paper. It can be applied in situations when it is necessary to make the promptly and reasonably decision on the communication satellites choice in conditions of limited time for decision-making. The stated choice method differs from the known ones by taking into account the overall characteristics of the satellite, which significantly affect the quality of its intended*

tasks. The estimated characteristics in the work are the availability, bandwidth, noise immunity, reliability and controlability of the satellite.. The paper also presents the approbation results of the propsed method to determine the most effective communication satellite in a geostationary orbital grouping belonging to the Russian Federation, involved in the organization of information exchange of mobile subscribers on the territory of the state.

Keywords: *communication satellite's target efficiency, the communication satellite's target efficiency estimating method, the estimating bandwidth availability, the communication satellite's noise immunity, the communication satellite's reliability, the communication satellite's manageability.*

References

1. Petukhov G.B. *Foundations of the Theory of the Effectiveness of Purposeful Processes. Part 1. Methodology, Methods, Models*. Moscow: USSR Ministry of Defense Publ.; 1989. 660 p. (in Russ.)
2. Fomin L.A., Budko P.A. *Efficiency and Quality of Infocommunication Systems. Optimization Methods*. Moscow: Fiziko-matematicheskaya literatura Publ.; 2008. 296 p. (in Russ.)
3. Solodkaya M.S. Reliability, efficiency, quality of control systems. *QUALITY.EUP.RU*. Available from: https://www.quality.eup.ru/MATERIALY10/qsm.htm [Accessed 13 August 2020)
4. Manuilov Yu.S., Pavlov A.N., Novikov E.A. System Analysis and Organization of Automated Control of Spacecraft. St. Petersburg: Military Space Academy of A.F. Mozhaysky Publ.; 2010. 266 p. (in Russ.)
5. GOST P 53801-2010 *Federal communication. Terms and determinations.* Moscow: Standartinform Publ.; 2011. (In Russ.)
6. Zhukov S., Kovalsky A., Kvasov M., Mitryaev G. Operational distribution of the radio resource of satellite communication system for the purpose of ensuring management of spacecrafts. *Trudy nauchno-issledovatelskogo instituta radio*. 2017;2:29–36. (in Russ.)
7. Zinnurov S., Kovalsky A., Mitryaev G. Task`s Solution of Satellite Communication System`s Optimal Radio Resource Planning for the Sessions of Space Appliance`s Orbital Group Managing. *Proc. of Telecom. Universities*. 20184(1):67–74. DOI:10.31854/1813-324x-2018-1-67-74
8. Parshutkin A.V., Maslakov P.A. Study of the Noise Immunity of Modern Standards of Satellite Communications to the Impact of Non-Stationary Interference. *SPIIRAS Proceedings*. 2017;4(53):159–177. (in Russ.) DOI:10.15622/sp.53.8
9. GOST P 56526-2015. *Reliability and safety requirements for space systems, complexes and unmanned spacecraft of unique (small series) production with long life of active operation*. Moscow: Standartinform Publ.; 2016. (in Russ.)
10. GOST P 53111-2008. *Stability of Functioning of a Public Communication Network. Requirements and check methods*. Moscow: Standartinform Publ.; 2019. (in Russ.)
11. GOST P 58625-2019 *Space systems and complexes. Maintainability analysis. General requirements.* Moscow: Standartinform Publ.; 2020. (in Russ.)
12. GOST P 58628-2019 *Space systems and complexes. Availability analysis. General requirements*. Moscow: Standartinform Publ.; 2020. (in Russ.)
13. *On-Board Spacecraft Control Systems* / Ed. A.S. Syrov. Moscow: MAI-PRINT Publ.; 2010. 304 p. (in Russ.)
14. Novikov E.A. *Flexible Technological Strategies for Managing the Target Application of the Space Navigation System*. St. Petersburg: Military Space Academy of A.F. Mozhaysky Publ.; 2009. 175 p. (in Russ.)
15. Manuilov Yu.S., Ptushkin A.I., Starodubov V.A. *Methodological Foundations for the Application of Flexible Spacecraft Control Strategies*. St. Petersburg: Ministry of Defense of the Russian Federation Publ.; 2002. 102 p. (in Russ.)
16. Manuilov Yu.S., Kalinin V.N., Goncharevsky V.S., et al. *Control of Spacecrafts and Means of Ground Control Complex.* St. Petersburg: Military Space Academy of A.F. Mozhaysky Publ.; 2010. 609 p. (in Russ.)

Сведения об авторах:

АБАЗИНА Евгения Сергеевна — кандидат технических наук, преподаватель кафедры сетей и систем связи космических комплексов Военно-космической академии имени А.Ф. Можайского, vka@mil.ru
https://orcid.org/0000-0001-5944-0316

КОВАЛЬСКИЙ Александр Александрович — кандидат технических наук, докторант военного института (научно-исследовательского) Военно-космической академии имени А.Ф. Можайского, vka@mil.ru
https://orcid.org/0000-0002-6878-5858

ПИТРИН Алексей Владимирович — адъюнкт военного института (научно-исследовательского) Военно-космической академии имени А.Ф. Можайского, vka@mil.ru

УДК 623.20 DOI:10.31854/1813-324X-2020-6-3-17-27

Модель информационных контактов устройств телекоммуникаций информационно-телекоммуникационной системы специального назначения со средствами наблюдения и воздействия противостоящей стороны

Р.Л. Михайлов[1*]

[1]Военный университет радиоэлектроники,
Череповец, 162622, Российская Федерация
[*]Адрес для переписки: cvviur6@mil.ru

Информация о статье
Поступила в редакцию 04.07.2020
Принята к публикации 12.08.2020

Ссылка для цитирования: Михайлов Р.Л. Модель информационных контактов устройств телекоммуникаций информационно-телекоммуникационной системы специального назначения со средствами наблюдения и воздействия противостоящей стороны // Труды учебных заведений связи. 2020. Т. 6. № 3. С. 17–27. DOI:10.31854/1813-324X-2020-6-3-17-27

Аннотация: *В настоящее время информационные конфликты являются неотъемлемой частью конфликтов между различными организационно-техническими системами во многих сферах деятельности. Решение задачи достижения информационного превосходства в информационном конфликте в специальной сфере возлагается на информационно-телекоммуникационные системы специального назначения (ИТКС СН). В соответствии с авторским подходом, такие ИТКС СН декомпозируются на телекоммуникационную и информационную составляющие, а также подсистемы наблюдения и воздействия. При этом для рассматриваемого информационного конфликта характерно наличие двух противоборствующих между собой ИТКС СН, которые соответствуют двум различным сторонам конфликта. В каждой из ИТКС СН средства подсистемы наблюдения призваны обеспечить систему управления необходимым объемом оперативной информации для принятия решений, а средства в составе подсистемы воздействия – затруднить процесс сбора информации со стороны противостоящей ИТКС СН. Конфликтное взаимодействие между ИТКС может быть представлено в форме информационных контактов средств наблюдения и воздействия с устройствами телекоммуникаций в составе противостоящей ИТКС СН. В связи с этим, формализация этих информационных контактов и анализ потенциального вклада указанных подсистем в достижение информационного превосходства в информационном конфликте является актуальной научной задачей. Целью работы является разработка модели информационных контактов и оценивание эффективности различных вариантов распределения устройств телекоммуникаций между подсистемами наблюдения и воздействия противостоящей стороны с позиции достижения информационного превосходства в информационном конфликте.*

Ключевые слова: *информационно-телекоммуникационная система специального назначения, радиомониторинг, радиоэлектронное подавление, информационный конфликт, устройство телекоммуникаций.*

Введение

Конец XX в. характеризовался масштабными изменениями в жизненном укладе – колоссальные достижения в различных отраслях науки и техники привели к глубоким изменениям в области глобализации общества. Важнейшей движущей силой этих процессов стала информатизация – глубокое проникновение информационных и коммуникационных технологий во все сферы жизни и деятельности человека. Тотальная информатизация общества привела к тому, что среди совокупности сфер жизнедеятельности современного общества на первое место вышла принципиально новая сфера – информационная. Этой сфере жизнедеятельности свойственны как новый ресурс – информация, так и новые противоречия, вызванные борьбой за обладание этим ресурсом [1], что, в свою очередь, стало одной из причин

интенсификации процесса конвергенции информационных и телекоммуникационных технологий в различных областях и создания на их основе единых информационно-телекоммуникационных систем (ИТКС) [2].

Под ИТКС в работе понимаются территориально распределенные комплексы средств управления и обработки информации (далее – информационных устройств) и средств связи (далее – устройств телекоммуникаций (УТ)), средства наблюдения и воздействия в составе соответствующих подсистем, а также соединяющие их каналы радиосвязи, обеспечивающие формирование, передачу, прием, хранение, поиск, отображение и обработку информации по заданным алгоритмам и программам и предназначенные для предоставления человеку и его организациям различных информационных и коммуникационных продуктов и услуг [2].

Создание подобных ИТКС стало следствием стремления обеспечить органы управления необходимым для принятия решений объемом информации, а также требуемой совокупностью телекоммуникационных услуг для повышения устойчивости реализации ими своих управленческих функций. Ограниченность информации как общего ресурса конфликтующих организационно-технических систем привело к появлению нового типа конфликтов с участием ИТКС – информационных конфликтов.

Информационный конфликт ИТКС – процесс столкновения ИТКС в различных сферах деятельности на этапе сбора и обработки данных о состоянии, намерениях и действиях противостоящей стороны, каждая из которых стремится к упреждающему по отношению к противостоящей стороне принятию управленческих решений и предпринимает определенные действия по снижению возможностей противостоящих средств сбора и обработки данных [3, 4].

В соответствии с [5] выделяют антагонистические и нестрогие информационные конфликты, при этом последние могут быть переведены к классу «коалиция» или «кооперация», так как противоречия могут быть устранены на основе реализации принципа согласованного оптиума [6–8] и в дальнейшем не рассматриваются. Антагонистические информационные конфликты с обратно пропорциональным или непропорциональным изменением показателей эффективности в большей степени характерны для ИТКС специального назначения (ИТКС СН), под которыми в соответствии с [9] в работе понимаются ИТКС, используемые для нужд органов государственной власти, обороны страны, безопасности государства и обеспечения правопорядка. Целью противостоящих в информационном конфликте ИТКС СН является достижение информационного превосходства, т. е. способности осуществлять непрерывный сбор сведений о противостоящей стороне, их обработку, распределение потока достоверной информации в интересах применения основных (базовых) средств, а также способность обеспечить упреждение выполнения аналогичных действий противостоящей стороной [3, 4]. В качестве примера, при рассмотрении ИТКС военного назначения, под основными (базовыми) средствами, как правило, понимаются силы и средства, непосредственно ведущие военное противоборство – воинские части и подразделения, ударные системы, комплексы огневого поражения и т. д.

Задачи сбора данных о состоянии, намерениях и действиях противостоящей стороны возлагаются на средства наблюдения в составе соответствующей подсистемы ИТКС СН, а затруднение выполнения аналогичных действий противостоящей стороной – на средства из состава подсистемы воздействия ИТКС СН. Введено допущение, что информационный обмен между УТ в составе ИТКС СН осуществляется по каналам радиосвязи, которые доступны средствам наблюдения и воздействия ИТКС другой стороны. При таком допущении средства наблюдения можно интерпретировать как средства радиомониторинга, а средства воздействия – как средства радиоэлектронного подавления (РЭП).

Решение задач сбора данных о состоянии, намерениях и действиях противостоящей стороны и затруднение выполнения аналогичных действий противостоящей стороной достигается посредством информационных контактов средств наблюдения и воздействия с УТ из состава телекоммуникационной подсистемы ИТКС СН противостоящей стороны.

Под информационным контактом средства наблюдения с УТ противостоящей стороны понимается процесс вскрытия местоположения УТ, режимов его работы и содержания радиообмена между абонентами телекоммуникационной подсистемы ИТКС СН противостоящей стороны [5, 10–12]. Результатом подобного информационного контакта является собранный объем оперативной информации, которая в дальнейшем обрабатывается в информационной подсистеме ИТКС СН и используется органами управления при принятии решений о применении основных (базовых) средств.

Под информационным контактом средства воздействия с УТ противостоящей стороны понимается процесс постановки помехи в канале радиосвязи, который соединяет данное устройство с другими УТ в составе телекоммуникационной подсистемы ИТКС СН [5, 13–17]. Результатом подобного информационного контакта является потеря противостоящей стороной некоторого объема оперативной информации, передаваемой во время контакта в канале радиосвязи, которая необходима органам

управления при принятии решений о применении основных (базовых) средств.

Таким образом, совокупность УТ из состава телекоммуникационной подсистемы ИТКС СН выступает как общий ресурс подсистем наблюдения и воздействия противостоящей стороны для реализации информационных контактов, при этом информационная подсистема в составе ИТКС СН выступает как орган, координирующий данные подсистемы, задачами которого является распределение между ними УТ телекоммуникационной подсистемы в составе ИТКС СН противостоящей стороны достижения информационного превосходства в информационном конфликте.

Концептуальная схема информационного конфликта ИТКС СН представлена на рисунке 1.

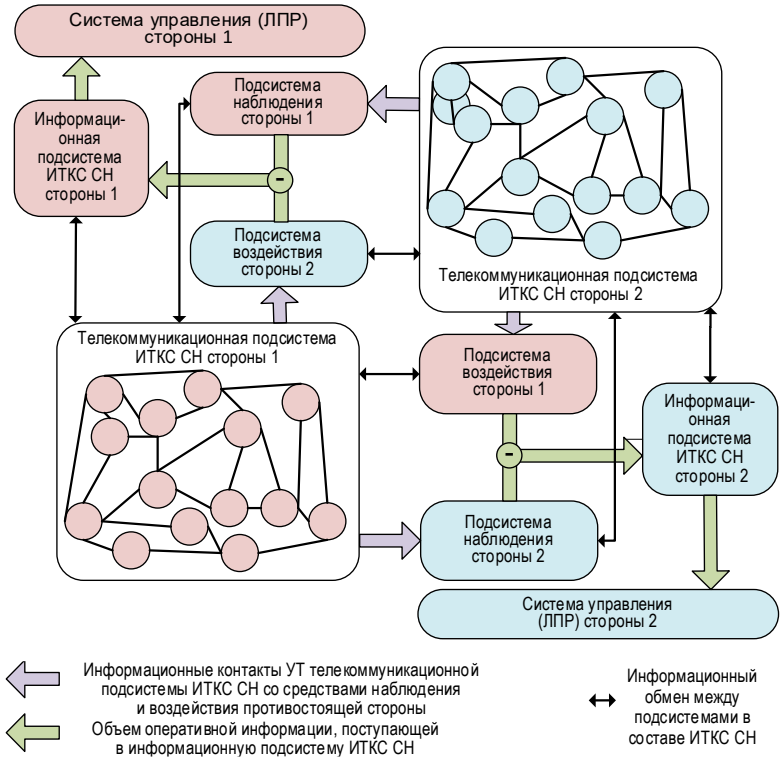


Рис. 1. Концептуальная модель информационного конфликта ИТКС СН

Fig. 1. The Conceptual Model of Special Purposes Infocommunication Systems Information Conflict

На рисунке 1 показано, что объем оперативной информации, поступающий в информационную подсистему ИТКС СН, определяется разностью между объемом информации, полученной средствами подсистемы наблюдения посредством информационных контактов с УТ телекоммуникационной подсистемы ИТКС СН противостоящей стороны, и объемом информации, утерянной в телекоммуникационной подсистеме ИТКС СН вследствие информационных контактов УТ со средствами подсистемы воздействия противостоящей стороны. С учетом ограниченности числа УТ телекоммуникационной подсистемы ИТКС СН противостоящей актуальным является оценивание эффективности их информационных контактов со средствами наблюдения и воздействия в интересах достижения их информационного превосходства в информационном конфликте, что позволит в рамках дальнейших исследований определить порядок оптимального распределения УТ телекоммуникационной подсистемы между подсистемами наблюдения и воздействия в рамках их координации со стороны информационной подсистемы ИТКС СН.

Постановка задачи

Анализ исследований информационных конфликтов, проведенный автором в работе [3], показал, что вопросами оценивания эффективности информационных контактов УТ ИТКС СН со средствами наблюдения противостоящей стороны занимались А.И. Куприянов, А.В. Сахаров, В.Г. Радзиевский, А.А. Сирота, С.В. Дворников и другие ученые. В их работах подробно рассматриваются пути повышения объема добываемой средствами наблюдения оперативной информации. Вместе с тем, оценивание вклада средств наблюдения в достижение информационного превосходства в информационном конфликте ИТКС СН в данных исследованиях не производилось.

Информационные контакты УТ телекоммуникационной подсистемы ИТКС СН со средствами воздействия противостоящей стороны широко исследованы в работах Ю.Л. Козирацкого, А.И. Палия, В.И. Владимирова, С.М. Одоевского и ряда других ученых. В данных работах рассматриваются, как правило, эффекты от информационных контактов УТ телекоммуникационной подсистемы ИТКС СН со

средствами воздействия противостоящей стороны на одном уровне – физическом (сигнальном), при этом не оценивается вклад подсистемы воздействия в достижение превосходства в информационном конфликте.

Кроме того, в работах А.М. Чуднова, А.В. Боговика, В.В. Игнатова, В.А. Цимбала, Ю.И. Стародубцева и целого ряда других ученых исследованы вопросы скрытности, помехоустойчивости и, как следствие, устойчивости телекоммуникационной подсистемы ИТКС СН. Однако в данных работах основное внимание уделяется именно вопросам повышения вышеуказанных характеристик телекоммуникационной подсистемы ИТКС СН, в то время как оценивание эффективности функционирования средств наблюдения и воздействия противостоящей стороны не осуществляется.

Таким образом, в настоящее время ведутся активные исследования информационных контактов УТ телекоммуникационной подсистемы ИТКС СН со средствами наблюдения и воздействия противостоящей стороны, однако в большинстве работ формализованы дуэльные ситуации либо «УТ – средство наблюдения», либо «УТ – средство воздействия», при этом не осуществляется оценивание эффективности комплексного применения подсистем наблюдения и воздействия в информационном конфликте ИТКС СН. Кроме того, не учитывается возможность координации подсистем наблюдения и воздействия за счет смены стратегии распределения УТ телекоммуникационной подсистемы ИТКС СН между средствами наблюдения и воздействия.

Вместе с тем, схожая с авторской постановка задачи исследования присутствует в работах Ю.И. Стародубцева, В.В. Бухарина, С.С. Семенова [18–21], С.И. Макаренко [5, 22–24], Ю.Л. Козирацкого [4, 25–28], В.И. Владимирова, В.П. Лихачева и В.М. Шляхина [29–32]. При этом в работах [18–21] обосновано расширение сферы ведения информационного конфликта и выход его за пределы традиционной сферы – радиосвязи, в глобальное телекоммуникационное пространство. При общих с автором подходах к структурам противостоящих ИТКС СН и функциям подсистем в их составе, имеют место существенные различия, т. к. телекоммуникационная подсистема ИТКС СН в соответствии с указанными работами представляет собой общую для обеих сторон информационного конфликта совокупность УТ и каналов связи различного рода между ними, что не позволяет классифицировать подобный информационный конфликт ИТКС СН как строго антагонистический.

В работах [5, 22–24] исследованы информационные контакты УТ телекоммуникационной подсистемы СН со средствами наблюдения и воздействия противостоящей стороны. Однако телекоммуникационные подсистемы СН рассматриваются как самостоятельные организационно-технические системы, и, таким образом, сложный многокомпонентный характер ИТКС СН не учитывается. Кроме того, в этих работах информационное превосходство в информационном конфликте «телекоммуникационные подсистемы СН – средства наблюдения и средства воздействия» формализуется показателями устойчивости ТКС СН, что, на взгляд автора, является не совсем корректным.

Работы [4, 25–28] посвящены исследованию информационных конфликтов телекоммуникационной подсистемы СН со средствами воздействия, представленными, главным образом, средствами РЭП. В соответствии с принятым в этих работах подходом, средствам наблюдения отводится лишь функции целеуказания в интересах эффективного применения средств воздействия, а также контроля результатов их применения. Таким образом, вклад подсистемы наблюдения в достижение информационного превосходства в информационном конфликте ИТКС СН не оценивается. Схожие подходы характерны и для работ [29–32], но в них введено понятие информационного ущерба телекоммуникационной подсистемы ИТКС СН вследствие информационных контактов УТ телекоммуникационной подсистемы со средствами воздействия (средствами РЭП) противостоящей стороны, а также оценивается влияние этого ущерба на обоснованность и своевременность принятия решений в информационной подсистеме ИТКС СН. Информационный ущерб в указанных работах определяется как отношение объема информации, переданной телекоммуникационной подсистемой СН в условиях воздействия средств РЭП на линии радиосвязи между УТ и объема информации, соответствующей требованиям по достоверности ее передачи в линиях радиосвязи. Вместе с тем, оценивание обоснованности принимаемых решений о применении основных (базовых) средств осуществляется для той ИТКС СН, УТ в составе телекоммуникационной подсистемы которой участвуют в информационных контактах со средствами воздействия противостоящей стороны, а информационный ущерб определяется как потеря объема передаваемой информации, необходимой для принятия решений в информационной подсистеме ИТКС СН. Вследствие этого, в явном виде не учитывается вклад подсистемы наблюдения в увеличение объема оперативной информации «своей» ИТКС СН, добываемой вследствие информационных контактов с УТ телекоммуникационной подсистемы, противостоящей ИТКС СН. В то же время сами эти подходы к формализации требований к обоснованности принимаемых решений, на взгляд автора, являются весьма перспективными и именно они легли в основу данного исследования.

Таким образом, отличительной особенностью авторского подхода к формализации информационных контактов УТ телекоммуникационной под-

системы ИТКС СН со средствами наблюдения и воздействия противостоящей стороны является наличие элементов оценивания вклада подсистем наблюдения и воздействия в достижение информационного превосходства в конфликте между двумя ИТКС СН.

Формализация модели информационных контактов устройств телекоммуникации информационно-телекоммуникационной системы специального назначения со средствами наблюдения и воздействия противостоящей стороны

Пусть имеются две ИТКС СН, которые соответствуют стороне 1 и стороне 2 информационного конфликта. Рассмотрим информационный конфликт с позиции стороны 1 и для формализации модели с позиции информационных контактов УТ телекоммуникационной подсистемы ИТКС СН стороны 2 со средствами наблюдения и воздействия стороны 1 введем следующие сокращения:

$V_{\text{пост}}$ – объем постоянной информации стороны 1, используемой при принятии решения;

$V_{\text{опер}}$ – объем оперативной информации, полученной ИТКС СН стороны 1 вследствие информационного контакта УТ ИТКС СН стороны 2 со средством наблюдения стороны 1;

$V_{\text{в}}$ – объем информации стороны 2 не переданной вследствие информационного контакта УТ телекоммуникационной подсистемы ИТКС СН со средством воздействия стороны 1;

$\gamma_{\text{в}}$ – информационный ущерб ИТКС СН стороны 2 при информационном контакте УТ телекоммуникационной подсистемы ИТКС СН стороны 2 со средством воздействия ИТКС СН стороны 1;

$P_{\text{ош}}^{\text{тр}}$ – требуемая достоверность передачи сообщения в каналах радиосвязи телекоммуникационной подсистемы ИТКС СН стороны 2;

$P_{\text{с}}(L)$ – вероятность достоверной передачи сообщения в телекоммуникационной подсистеме стороны 2 за L повторов;

L – количество повторов передачи сообщения;

$P_{\text{с}}(1)$ – вероятность передачи сообщения в телекоммуникационной подсистеме стороны 2 за один сеанс;

$\xi(\gamma_{\text{в}})$ – относительное время задержки передачи сообщения вследствие информационного ущерба $\gamma_{\text{в}}$ от информационного контакта УТ телекоммуникационной подсистемы ИТКС СН стороны 2 со средством воздействия стороны 1;

$\Delta T_{\text{реш}_{\text{в}}}$ – выигрыш во времени принятия решения стороны 1 вследствие информационного контакта УТ телекоммуникационной подсистемы ИТКС СН стороны 2 со средством воздействия стороны 2;

V – объем информации стороны 2, передаваемый через УТ ИТКС СН стороны 2 в ходе информационного контакта этого УТ со средством наблюдения или средством воздействия стороны 1;

$T_{\text{с}}$ – длительность передачи сообщения в телекоммуникационной подсистеме ИТКС СН стороны 2;

$T_{\text{зс}}$ – время задержки передачи сообщения в ТКС ИТКС СН стороны 2, вследствие информационного контакта УТ ТКС со средством воздействия стороны 1;

$\gamma_{\text{н}}$ – информационный ущерб ИТКС СН стороны 2 при информационном контакте УТ телекоммуникационной подсистемы ИТКС СН стороны 2 со средством наблюдения ИТКС СН стороны 1;

$\boldsymbol{N}_{\text{вар}}$ – число вариантов принятия решения стороной 1;

$T_{\text{вар}}$ – время, необходимое для отработки одного варианта принятия решения стороной 1;

$\Delta T_{\text{реш}_{\text{н}}}$ – выигрыш во времени принятия решения стороны 1 вследствие информационного контакта УТ телекоммуникационной подсистемы ИТКС СН стороны 2 со средством наблюдения стороны 2;

Q – множество УТ в составе телекоммуникационной подсистемы ИТКС СН стороны 2;

$\boldsymbol{q}$ – УТ в составе телекоммуникационной подсистемы ИТКС СН стороны 2, $\boldsymbol{q} \in Q$.

На рисунке 2 приведена схема модели информационных контактов УТ ИТКС СН со средствами наблюдения и воздействия противостоящей стороны.

При информационном контакте УТ телекоммуникационной подсистемы ИТКС СН стороны 2 со средством воздействия ИТКС СН стороны 1 информационный ущерб ИТКС СН стороны 2 $\gamma_{\text{в}}$ определяется отношением объема информации стороны 2, не переданной вследствие этого информационного контакта $V_{\text{в}}$, к общему объему информации стороны 2, передаваемой через УТ ИТКС СН стороны 2 в ходе этого информационного контакта V:

$$\gamma_{\text{в}} = \frac{V_{\text{в}}}{V}. \qquad (1)$$

При этом в современных телекоммуникационных подсистемах ИТКС СН для повышения достоверности передачи сообщений в условиях воздействия помех на каналы радиосвязи используется их повторная передача, что позволяет обеспечить выполнение условия:

$$P_{\text{ош}}^{\text{тр}} = 1 - P_{\text{с}}(L), \qquad (2)$$

где $P_{\text{ош}}^{\text{тр}}$ – требуемая достоверность передачи сообщения в каналах радиосвязи телекоммуникационной подсистемы ИТКС СН; $P_{\text{с}}(L)$ – вероятность достоверной передачи сообщения за L повторов; L – количество повторов передачи сообщения.

Вероятность достоверной передачи сообщения $P_{\text{с}}(L)$ определяется выражением:

$$P_{\text{с}}(L) = 1 - [1 - P_{\text{с}}(1)]^{L}, \qquad (3)$$

где $P_{\text{с}}(1)$ – вероятность передачи сообщения за один сеанс; L – количество повторов передачи сообщения.

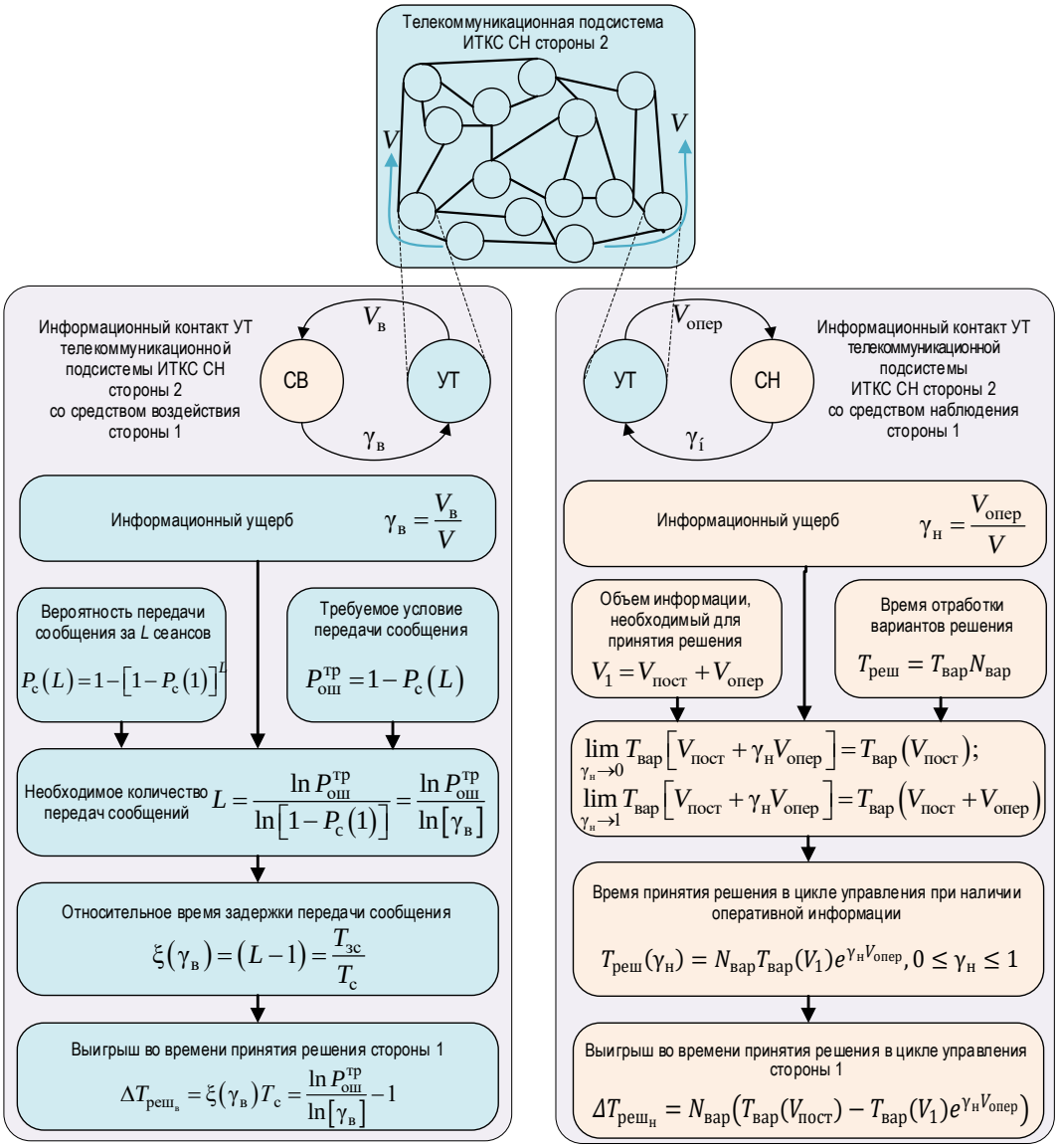


Рис. 2. Схема модели информационных контактов УТ ИТКС СН со средствами наблюдения и воздействия

Fig. 2. Scheme of the Model of Information Contacts Between Telecommunication Devices of Special Purpose Infocommunication System and Means of Monitoring and Impact

Таким образом, требуемое количество повторов передачи сообщения L, необходимое для выполнения требования по достоверности его передачи при допущении о равновероятном распределении состояний символов в сообщении и одинаковой вероятности ложных переходов за один сеанс [29], определяется выражением:

$$L = \frac{\ln P_{ош}^{тр}}{\ln[1 - P_с(1)]} = \frac{\ln P_{ош}^{тр}}{\ln[\gamma_в]}. \quad (4)$$

Тогда относительное время задержки передачи сообщения вследствие информационного ущерба $\gamma_в$ от информационного контакта УТ телекоммуникационной подсистемы ИТКС СН стороны 2 со средством воздействия стороны 1 $\xi(\gamma_в)$ определяется из выражения:

$$\xi(\gamma_в) = (L - 1) = \frac{T_{зс}}{T_с}, \quad (5)$$

где $T_{зс}$ – время задержки передачи сообщения в телекоммуникационной подсистеме ИТКС СН стороны 2 вследствие информационного контакта УТ телекоммуникационной подсистемы со средством воздействия стороны 1; $T_с$ – длительность передачи сообщения в телекоммуникационной подсистеме ИТКС СН стороны 2.

Тогда выигрыш во времени принятия решения стороны 1 вследствие информационного контакта УТ телекоммуникационной подсистемы ИТКС СН стороны 2 со средством воздействия стороны 2 $\Delta T_{реш_в}$ составит:

$$\Delta T_{реш_в} = \xi(\gamma_в) Tc = \frac{\ln P_{ош}^{тр}}{\ln[\gamma_в]} - 1. \quad (6)$$

Зависимость выигрыша во времени принятия решения стороны 1 вследствие информационного контакта УТ телекоммуникационной подсистемы ИТКС СН стороны 2 со средством воздействия стороны 1 $\Delta T_{\text{реш}_\text{в}}$ от величины информационного ущерба $\gamma_\text{в}$ при различных значениях требуемой достоверности передачи сообщения в каналах радиосвязи телекоммуникационной подсистемы ИТКС СН стороны 2 $P_{\text{ош}}^{\text{тр}}$ и длительности передачи сообщения в телекоммуникационной подсистеме ИТКС СН стороны 2 $T_\text{с} = 1$ с представлена на рисунке 3.

Анализ представленной зависимости позволяет сделать вывод о том, что информационный контакт УТ телекоммуникационной подсистемы ИТКС СН стороны 2 со средством воздействия ИТКС СН стороны 1 может привести к задержке в передаче информации, необходимой стороне 2 для принятия решения, на длительность до трех порядков относительно времени передачи сообщения в телекоммуникационной подсистеме ИТКС СН стороны 2 при отсутствии помех в каналах радиосвязи T_c. Другим неоднозначным выводом является то, что в интересах снижения выигрыша ИТКС СН стороны 1 во времени принятия решения $\Delta T_{\text{реш\_в}}$ стороне 2 необходимо снизить требования по достоверности передачи информации в своих каналах радиосвязи $P_{\text{ош}}^{\text{тр}}$.

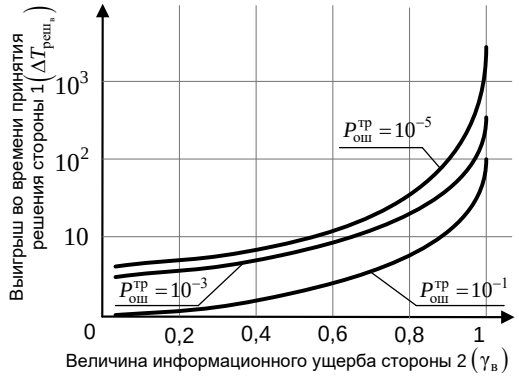


Рис. 3. Зависимость выигрыша во времени принятия решения ИТКС СН стороны 1 от информационного ущерба вследствие информационного контакта УТ телекоммуникационной подсистемы ИТКС СН стороны 2 со средством воздействия стороны 1

Fig. 3. The Correlation of the Gain in the Decision Time of the Side 1 Special Purpose Infocommunication System and the Information Damage Due to the Information Contact Between Telecommunication Devices and Side 2 Mean of Impact

При информационном контакте УТ телекоммуникационной подсистемы ИТКС СН стороны 2 со средством наблюдения ИТКС СН стороны 1 (рисунок 2) информационный ущерб ИТКС СН стороны 2 $\gamma_\text{н}$ определяется отношением объема полученной ИТКС СН стороны 1 вследствие этого информационного контакта оперативной информации $V_{\text{опер}}$ к общему объему информации стороны 2, передаваемой через УТ ИТКС СН стороны 2 в ходе этого информационного контакта V:

$$\gamma_\text{н} = \frac{V_{\text{опер}}}{V}. \qquad (7)$$

Указанный объем оперативной информации $V_{\text{опер}}$, складываясь с объемом постоянной информации ИТКС СН стороны 1 $V_{\text{пост}}$, образует общий объем используемой в процессе принятия решения стороной 1 информации V_1:

$$V_1 = V_{\text{пост}} + V_{\text{опер}}, \qquad (8)$$

где $V_{\text{пост}}$ – объем постоянной информации стороны 1, используемой при принятии решения; $V_{\text{опер}}$ – объем оперативной информации, полученной ИТКС СН стороны 1 вследствие информационного контакта УТ ИТКС СН стороны 2 со средством наблюдения стороны 1.

Как показано в работе [29], для принятия решения системе управления (ЛПР) стороны 1 необходимо рассмотреть определенное количество вариантов $\boldsymbol{N}_{\text{вар}}$, при этом отработка каждого варианта решения требует определенных временных затрат $T_{\text{вар}}$. Таким образом, временные затраты на принятие решения $T_{\text{реш}}$ стороной 1 определяются выражением:

$$T_{\text{реш}} = T_{\text{вар}} N_{\text{вар}}. \qquad (9)$$

При этом временные затраты на отработку одного варианта решения $T_{\text{вар}}$ при различных значениях информационного ущерба стороны 2 $\gamma_\text{н}$ составляют:

$$\begin{aligned} &\lim_{\gamma_\text{н}\to 0} T_{\text{вар}}\left(V_{\text{пост}} + \gamma_\text{н} V_{\text{опер}}\right) = T_{\text{вар}}(V_{\text{пост}}); \\ &\lim_{\gamma_\text{н}\to 1} T_{\text{вар}}\left(V_{\text{пост}} + \gamma_\text{н} V_{\text{опер}}\right) = T_{\text{вар}}\left(V_{\text{пост}} + V_{\text{опер}}\right). \end{aligned} \qquad (10)$$

Таким образом, увеличение значения времени принятия решения $T_{\text{реш}}$ стороной 1 нелинейно связано с ростом информационного ущерба стороны 2 $\gamma_\text{н}$. В работе [29] обоснован экспоненциальный характер этой зависимости, тогда выражение (9) с учетом (8) примет вид:

$$T_{\text{реш}}(\gamma_\text{н}) = N_{\text{вар}} T_{\text{вар}}(V_1) e^{\gamma_\text{н} V_{\text{опер}}}, 0 \le \gamma_\text{н} \le 1. \qquad (11)$$

Выигрыш во времени принятия решения стороны 1 $\Delta T_{\text{реш}_\text{н}}$ определяется разницей между принятием решения в условиях наличия объема оперативной информации $V_{\text{опер}}$ и ее отсутствием, т. е. принятием решения только на основе имеющегося объема постоянной информации $V_{\text{пост}}$:

$$\Delta T_{\text{реш}_\text{н}} = N_{\text{вар}}\left(T_{\text{вар}}(V_{\text{пост}}) - T_{\text{вар}}(V_1) e^{\gamma_\text{н} V_{\text{опер}}}\right). \qquad (12)$$

Зависимость выигрыша во времени принятия решения стороны 1 вследствие информационного контакта УТ телекоммуникационной подсистемы ИТКС СН стороны 2 со средством наблюдения стороны 1 $\Delta T_{\text{реш}_\text{н}}$ от величины информационного ущерба стороне 2 $\gamma_\text{н}$ в относительной шкале отношения временных затрат на принятия решения в условиях отсутствия объема оперативной информации $V_{\text{опер}}$ и его наличия $\frac{T_{\text{вар}}(V_{\text{пост}})}{T_{\text{вар}}(V_1)}$, представлена на рисунке 4.

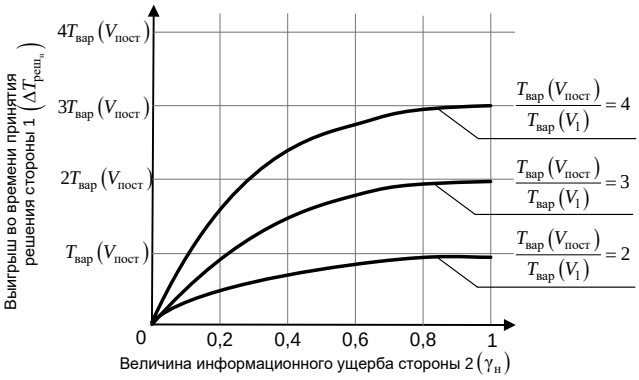


Рис. 4. Зависимость выигрыша во времени принятия решения ИТКС СН стороны 1 от информационного ущерба вследствие информационного контакта УТ телекоммуникационной подсистемы ИТКС СН стороны 2 со средством наблюдения стороны 1

Fig. 4. The Correlation of the Gain in the Decision Time of the Side 1 Special Purpose Infocommunication System and the Information Damage Due to the Information Contact Between Telecommunication Devices and Side 2 Mean of Monitoring

Анализ представленной зависимости позволяет сделать вывод о том, что информационный контакт УТ телекоммуникационной подсистемы ИТКС СН стороны 2 со средством наблюдения ИТКС СН стороны 1 приводит к снижению времени принятия решения стороной 1. Вместе с тем, следует указать на то, что в рамках предложенной модели не исследовались вопросы достоверности данных решений с позиции определения необходимого количество вариантов решения $N_{\text{вар}}$. Данная величина также оказывает влияние на значение время принятия решения в соответствии с выражениями (9) и (12), однако вопросы ее оптимизации относятся к сфере обработки оперативной информации в информационной подсистеме ИТКС СН, а не ее сбора.

Таким образом, каждое УТ ИТКС СН стороны 2 может быть представлено вектором $\mathbf{q} = \left(\Delta T_{\text{реш}_н}, \Delta T_{\text{реш}_в}\right), \boldsymbol{q} \in Q$, характеризующим его с позиции потенциального выигрыша во времени принятия решения стороной 1, что позволяет перейти к определению оптимального распределения этих УТ для информационных контактов со средствами наблюдения и воздействия.

Заключение

Представленная в работе модель позволяет формализовать вклад подсистем наблюдения и воздействия в составе ИТКС СН в достижение информационного превосходства в информационном конфликте. В работе рассматриваются, главным образом, телекоммуникационная подсистема СН, в которых в качестве среды распространения сигналов между УТ используется радиоэфир, однако предложенный подход с незначительными уточнениями может быть применен и к телекоммуникационной подсистеме с каналами связи другой физической природы (оптическими, проводными и т. д.). В рамках исследования модели выявлена зависимость времени принятия решения от объема собранной средствами наблюдения оперативной информации, а также от объема непереданной вследствие информационных контактов УТ ИТКС СН со средствами воздействия противостоящей стороны информации. Следует отметить, что выигрыши во времени принятия решения одной из сторон информационного конфликта вследствие информационных контактов средств наблюдения и воздействия в ее составе не являются строго пропорциональными, т. е. одно и то же УТ ИТКС СН противостоящей стороны может быть предпочтительным как более с позиции наблюдения за ним, так и с позиции воздействия на него. Более того, при распределении этих УТ необходимо учитывать двунаправленный характер информационного конфликта ИТКС СН, а, следовательно, и аналогичное распределение, которое производит противостоящая сторона. В этой связи, дальнейшим направлением исследований является разработка элементов научно-методического аппарата координации [33–36] подсистем наблюдения и воздействия посредством оптимального динамического распределения между ними УТ ИТКС СН противостоящей стороны для информационных контактов [8, 37]. Кроме того, при данном распределении УТ должны учитываться возможности противостоящей телекоммуникационной подсистеме ИТКС по динамической маршрутизации информационных потоков [38].

Список используемых источников

1. Макаренко С.И., Иванов М.С. Сетецентрическая война – принципы, технологии, примеры и перспективы: монография. СПб.: Наукоемкие технологии, 2018. 898 с.

2. Воробьев С.П., Давыдов А.Е., Ефимов В.В., Курносов В.И. Инфокоммуникационные сети: энциклопедия. Том 1: Инфокоммуникационные сети: классификация, структура, архитектура, жизненный цикл, технологии. СПб.: Наукоемкие технологии, 2019. 739 с.

3. Макаренко С. И., Михайлов Р.Л. Информационные конфликты – анализ работ и методологии исследований // Системы управления, связи и безопасности. 2016. № 3. С. 95–178. DOI:10.24411/2410-9916-2016-10304

4. Козирацкий Ю.Л., Будников С.А., Гревцев А.И., Иванцов А.В., Кильдюшевский В.М., Козирацкий А.Ю. и др. Модели информационного конфликта средств поиска и обнаружения: монография. М.: Радиотехника, 2013. 232 с.

5. Макаренко С.И. Модели системы связи в условиях преднамеренных дестабилизирующих воздействия и ведения разведки: монография. СПб.: Наукоемкие технологии, 2020. 337 с.

6. Владимиров В.И., Докторов А.Л., Елизаров Ф.В. и др. Электромагнитная совместимость радиоэлектронных средств и систем. М.: Радио и связь, 1985. 272 с.

7. Дружинин В.В., Конторов А.С., Конторов Д.С. Введение в теорию конфликта. М.: Радио и связь, 1989. 288 с.
8. Михайлов Р.Л., Ларичев А.В., Смыслова А.Л., Леонов П.Г. Модель распределения ресурсов в информационном конфликте организационно-технических систем // Вестник Череповецкого государственного университета. 2016. № 6. С. 24–29.
9. Федеральный закон Российской Федерации от 07.07.2003 № 126-ФЗ (ред. от 01.05.2016) «О связи».
10. Радзиевский В.Г., Сирота А.А. Теоретические основы радиоэлектронной разведки. М.: Радиотехника, 2004. 432 с.
11. Радзиевский В.Г., Сирота А.А. Информационное обеспечение радиоэлектронных систем в условиях конфликта. М.: ИПРЖР, 2001. 456 с.
12. Демин В.П., Куприянов А.И., Сахаров А.В. Радиоэлектронная разведка и радиомаскировка. М.: МАИ, 1997. 155 с.
13. Палий А.И. Радиоэлектронная борьба. М.: Воениздат, 1989. 350 с.
14. Перунов Ю.М., Фомичев К.И., Юдин Л.М. Радиоэлектронное подавление информационных каналов систем управления оружием. М.: Радиотехника, 2003. 416 с.
15. Куприянов А.И., Шустов Л.Н. Радиоэлектронная борьба. Основы теории. М.: Вузовская книга, 2011. 800 с.
16. Цветнов В.В., Демин В.П., Куприянов А. И. Радиоэлектронная борьба: радиомаскировка и помехозащита. М.: МАИ, 1999. 240 с.
17. Куприянов А. И. Радиоэлектронная борьба. М.: Вузовская книга, 2013. 360 с.
18. Стародубцев Ю.И., Бухарин В.В., Семенов С.С. Техносферная война // Военная мысль. 2012. № 7. С. 22–31.
19. Стародубцев Ю.И., Бухарин В.В., Семенов С.С. Техносферная война // Информационные системы и технологии. 2011. № 1. С. 80–85.
20. Стародубцев Ю.И., Семенов С.С., Бухарин В.В. Техносферная война // Армия и общество. 2010. № 4. С. 6–11.
21. Семенов С.С., Гусев А.П., Барботько Н.В. Оценка информационно-боевого потенциала сторон в техносферных конфликтах // Наукоемкие технологии в космических исследованиях Земли. 2013. Т. 5. № 6. С. 10–21.
22. Макаренко С.И. Динамическая модель двунаправленного информационного конфликта с учетом возможностей сторон по наблюдению, захвату и блокировке ресурса // Системы управления, связи и безопасности. 2017. № 1. С. 60–97.
23. Макаренко С.И. Динамическая модель системы связи в условиях функционально-разноуровневого информационного конфликта наблюдения и подавления // Системы управления, связи и безопасности. 2015. № 3. С. 122–185. DOI:10.24411/2410-9916-2015-10307
24. Макаренко С.И. Информационный конфликт системы связи с системой дестабилизирующих воздействий. Часть. I: Концептуальная модель конфликта с учетом ведения разведки, физического, радиоэлектронного и информационного поражения средств связи // Техника радиосвязи. 2020. Выпуск 2(45). С. 104–117. DOI:10.33286/2075-8693-2020-45-104-117
25. Козирацкий Ю.Л., Подлужный В.И., Паринов М.Л. Методический подход к построению вероятностной модели конфликта сложных систем // Вестник ВИРЭ. 2005. № 3. С. 4–16.
26. Козирацкий Ю.Л., Ерофеев А.Н., Соколовский С.П. Модель конфликтного взаимодействия «нарушитель – подсистема защиты информации автоматизированной системы управления» // Вестник Военного авиационного инженерного университета. 2012. № 1(15). С. 210–217.
27. Ухин А.Л., Козирацкий Ю.Л. Вероятностная модель конфликта радиоэлектронных систем управления и телекоммуникации в условиях деструктивных воздействий // Системы управления и информационные технологии. 2014. Т. 57. № 3-2. С. 281–286.
28. Козирацкий Ю.Л., Кущев С.С., Чернухо И.И., Донцов А.А. Модель конфликтного взаимодействия систем управления противоборствующих сторон в условиях преднамеренных помех // Радиотехника. 2012. № 5. С. 56–61.
29. Владимиров В.И., Лихачев В.П., Шляхин В.М. Антагонистический конфликт радиоэлектронных систем. М.: Радиотехника, 2004. 384 с.
30. Владимиров В.И., Владимиров И.В. Основы оценки конфликтно-устойчивых состояний организационно-технических систем (в информационных конфликтах). Воронеж: ВАИУ, 2008. 231 с.
31. Владимиров В.И. Принципы и аппарат системных исследований радиоэлектронного конфликта. Воронеж: ВВВИУРЭ, 1992. 153 с.
32. Владимиров В.И. Информационные основы радиоподавления линий радиосвязи в динамике радиоэлектронного конфликта. Воронеж: ВИРЭ, 2003. 276 с.
33. Михайлов Р.Л. Анализ научно-методического аппарата теории координации и его использования в различных областях исследований // Системы управления, связи и безопасности. 2016. № 4. С. 1–29.
34. Михайлов Р.Л. Анализ подходов к формализации показателя информационного превосходства на основе теории оценки и управления рисками // Системы управления, связи и безопасности. 2017. № 3. С. 98–118.
35. Михайлов Р.Л. Двухуровневая модель координации подсистем радиомониторинга и радиоэлектронной борьбы // Наукоемкие технологии в космических исследованиях Земли. 2018. Т. 10. № 2. С. 43–50. DOI:10.24411/2409-5419-2018-10040
36. Михайлов Р.Л. Модель динамической координации подсистем наблюдения и воздействия в информационном конфликте в виде иерархической дифференциальной игры трех лиц // Наукоемкие технологии. 2018. Т. 19. № 10. С. 44–51. DOI:10.18127/j19998465-201810-08
37. Михайлов Р.Л., Поляков С.Л. Модель оптимального распределения ресурсов и исследование стратегий действий сторон в ходе информационного конфликта // Системы управления, связи и безопасности. 2018. № 4. С. 323–344.
38. Макаренко С.И., Рюмшин К.Ю., Михайлов Р.Л. Модель функционирования объекта сети связи в условиях ограниченной надежности каналов связи // Информационные системы и технологии. 2014. № 6(86). С. 139–147.

* * *

The Model of Information Contacts Between Telecommunication Devices of Special Purpose Infocommunication System and Opposing Side Means of Monitoring and Impact

R. Mikhailov[1]

[1]Military University of Radioelectronics,
Cherepovets, 162622, Russian Federation

Article info
DOI:10.31854/1813-324X-2020-6-3-17-27
Received 4th July 2020
Accepted 12th August 2020

For citation: Mikhailov R. The Model of Information Contacts Between Telecommunication Devices of Special Purpose Infocommunication System and Opposing Side Means of Monitoring and Impact. *Proc. of Telecom. Universities.* 2020;6(3):17–27. (in Russ.) DOI:10.31854/ 1813-324X-2020-6-3-17-27

Abstract: *Currently, information conflicts are an integral part of conflicts between different organizational and technical systems in many areas of activity. The solution to the problem of achieving information superiority in an information conflict in the special sphere is assigned to special purpose info-telecommunication systems. According to the author's approach, they include telecommunications and information components, as well as monitoring and impact subsystems. At the same time, the information conflict under consideration is characterized by the presence of two opposing special purposes infocommunication systems, which correspond to two different sides of the conflict. In each infocommunication systems, the means of the monitoring subsystem are designed to provide control systems with the necessary amount of operational information for making decision, and the means as part of the influence subsystem are making it difficult to collect information from the opposing special purposes infocommunication system. Conflict interaction between infocommunication system can be represented in the form of information contacts by means of observation and influence with telecommunication devices as part of the opposing infocommunication systems. In this regard, the formalization of these information contacts and analyzing the potential contribution of these subsystems to the information superiority achievement in the information conflict is an urgent scientific task. The aim of the work is to develop a model of contacts and to evaluate the effectiveness of various options for the distribution of telecommunication devices between the the monitoring subsystems and the impact of the opposing side from the standpoint of achieving information superiority in an information conflict.*

Keywords: *special purpose infocommunication system, radiomonitoring, jamming, information conflict, telecommunication device.*

References

1. Makarenko S.I., Ivanov M.S. *Network-Centric War – Principles, Technologies, Examples and Prospects*. St. Petersburg: Naukoemkie Tekhnologii Publ.; 2018. 898 p. (in Russ.)
2. Vorob'ev S.P., Davydov A.E., Efimov V.V., *Kurnosov V.I. Infocommunication Networks. Encyclopedia. Vol. 1. Information and Communication Networks, Classification, Structure, Architecture, Life Cycle, Technology*. St. Petersburg: Naukoemkie tehnologii Publ.; 2019. 739 p. (in Russ.)
3. Makarenko S.I., Mikhailov R.L. Information Conflicts – Analysis of Papers and Research Methodology. *Systems of Control, Communication and Security*. 2016;3:95–178. (in Russ.) DOI:10.24411/2410-9916-2016-10304
4. Koziratskii Iu.L., Budnikov S.A., Grevtsev A.I., Ivantsov A.V., Kildyushevsky V.M., Koziratsky A.Yu., et al. *Model Information Conflict of Search and Discovery*. Moscow: Radiotekhnika Publ.; 2013. 232 p. (in Russ.)
5. Makarenko S.I. *Models of the Communication System under Conditions of Deliberate Destabilizing Influence and Conducting Monitoring*. St. Petersburg: Naukoemkie Tekhnologii Publ.; 2020. 337 p. (in Russ.)
6. Vladimirov V.I., Doktorov A.L., Elizarov F.V., et al. *Electromagnetic Compatibility of Radio Electronic Means and Systems*. Moscow: Radio i Sviaz Publ.; 1985. 272 p. (in Russ.)
7. Druzhinin V.V., Kontorov A.S., Kontorov D.S. *Introduction to the Theory of Conflict*. Moscow: Radio i Sviaz Publ.; 1989. 288 p. (in Russ.)
8. Mikhailov R.L., Larichev A.V., Smyslova A.L., Leonov P.G. Model of Resource Allocation in Information Conflict of Complicated Organizational and Technical Systems. *Cherepovets State University Bulletin*. 2016;6:24–29. (in Russ.)
9. *About the Connection*. Federal law of Russia. 2003. (in Russ.)

10. Radzievskii V.G., Sirota A.A. *The Theoretical Basis of Electronic Intelligence*. Moscow: Radiotekhnika Publ.; 2004. 432 p. (in Russ.)
11. Radzievskii V.G., Sirota A.A. *Information Support of Electronic Systems in Conflict*. Moscow: IPRZR Publ.; 2001. 456 p. (in Russ.)
12. Demin V.P., Kupriianov A.I., Sakharov A.V. Electronic Reconnaissance and Radioactive. Moscow: Moscow Aviation Institute Publ.; 1997. 155 p. (in Russ.)
13. Paliy A.I. *Electronic Warfare*. Moscow: Voenizdat Publ.; 1989. 350 p. (in Russ.)
14. Perunov Ju.M., Fomichev K.I., Iudin L.M. *Electronic Suppression of Information Channels of Weapon Control Systems*. Moscow: Radiotekhnika Publ.; 2003. 416 p. (in Russ.)
15. Kuprijanov A.I., Shustov L.N. *Electronic Warfare. Fundamentals of the Theory*. Moscow: Vuzovskaia Kniga Publ.; 2011. 800 p. (in Russ.)
16. Tsvetnov V.V., Demin V.P., Kupriianov A.I. *Electronic Warfare: Radioactive and Jamming Protection*. Moscow: Moscow Aviation Institute Publ.; 1999. 240 p. (in Russ.)
17. Kuprijanov A.I. *Electronic Warfare*. Moscow: Vuzovskaia Kniga Publ.; 2013. 360 p. (in Russ.)
18. Starodubtsev Ju.I., Bukharin V.V., Semenov S.S. Techno War. *Military Thought*. 2012;7:22–31. (in Russ.)
19. Starodubtsev Ju.I., Bukharin V.V., Semenov S.S. Technospherny war. *Informatsionnye sistemy i tekhnologii*. 2011;1: 80–85. (in Russ.)
20. Starodubtsev Ju.I., Bukharin V.V., Semenov S.S. Techno War. *Armiia i obshchestvo.* 2010;4:6–11. (in Russ.)
21. Semenov S.S., Gusev A.P., Barbotko N.V. Assessment Information the Combat Potential of the Parties in Technosphere Conflicts. *H&ES Researh*. 2013;5(6):10–21 (in Russ.)
22. Makarenko S.I. Dynamic Model of the Bi-directional Information Conflict to Take into Account Capabilities of Monitoring, Capturing and Locking of Information Resources. *Systems of Control, Communication and Security*. 2017;1:60–97. (in Russ.)
23. Makarenko S.I. Dynamic Model of Communication System in Conditions the Functional Multilevel Information Conflict of Monitoring and Suppression. *Systems of Control, Communication and Security*. 2015;3:122–185. (in Russ.) DOI:10.24411/2410-9916-2015-10307
24. Makarenko S.I. Information Conflict Between a Communication System and a System of Destabilizing Influences. Part 1. A Conceptual Conflict Model Taking into Account Signal and Communications Intelligence, Fire Weapons, Electronic and Microwave Warfare, Cyber-attacks. *Radio Communication Technology*. 2020;2(45):104–117. (in Russ.) DOI:10.33286/2075-8693-2020-45-104-117
25. Koziratskiy Ju.L., Podluzhnyi V.I., Parinov M.L. Methodical Approach to Constructing Probabilistic Models of Complex Conflict Systems. *Vestnik of Military Institute of Radioelectronics*. 2005;3:4–16. (in Russ.)
26. Koziratsky Ju.L., Erofeev A.N., Sokolovskii S.P. Model of Conflict Interaction "Violator – the Subsystem of Information Security of Automated Control Systems". *Vestnik Voennogo aviatsionnogo inzhenernogo universiteta*. 2012;15(1):210–217. (in Russ.)
27. Ukhin A.L., Koziratsky Ju.L. Probabilistic Model of Conflict Radio-Electronic Control Systems and Telecommunications in Terms of Destructive Impacts. *Sistemy upravleniia i informatsionnye tekhnologii*. 2014; 57(3-2):281–286. (in Russ.)
28. Koziratsky Ju.L., Kuschev S.S., Chernuho I.I., Dontsov A.A. Model of Disputed Interaction of Control Systems of the Contradictory Parties in the Conditions of Deliberate Hindrances. *Radiotekhnika*. 2012;5:56–61. (in Russ.)
29. Vladimirov V.I., Likhachev V.P., Shliakhin V.M. *Antagonistic Conflict of Radio-Electronic Systems*. Moscow: Radiotekhnika Publ.; 2004. 384 p. (in Russ.)
30. Vladimirov V.I., Vladimirov I.V. *Basis of Assessment of the Conflict-Stable States of Organizational and Technical Systems (in Information Conflicts)*. Voronezh: Military Aviation Engineering University Publ.; 2008. 231 p. (in Russ.)
31. Vladimirov V.I. *The Principles and Apparatus of the Electronic System Studies Conflict*. Voronezh: Voronezh Higher Military Engineering College of Radioelectronics Publ.; 1992. 153 p. (in Russ.)
32. Vladimirov V.I. *Information Basis of the Countermeasure of Radio Communications in the Dynamics of Electronic Conflict*. Voronezh: Military Engineering College of Radioelectronics Publ.; 2003. 276 p. (in Russ.)
33. Mikhailov R.L. An Analysis of the Scientific and Methodological Apparatus of Coordination Theory and its Use in Various Fields of Study. *Systems of Control, Communication and Security*. 2016;4:1–29. (in Russ.)
34. Mikhailov R.L. Analysis of Approaches to the Formalization of the Indicator of Information Superiority Based on the Theory of Assessment and Risk Management. *Systems of Control, Communication and Security*. 2017;3:98–118. (in Russ.)
35. Mihailov R. L. Two-level Model of Coordination of Subsystems of Radiomonitoring and Electronic Warfare. *H&ES Researh*. 2018;10(2):43–50. (in Russ.) DOI:10.24411/2409-5419-2018-10040
36. Mikhailov R.L. Model of Dynamic Coordination of Subsystems of Surveillance and Impact in the Information Conflict as a Hierarchical Differential Game of Three Sides. *Journal Science Intensive Technologies*. 2018;19(10):44–51. (in Russ.) DOI:10.18127/j19998465-201810-08
37. Mikhailov R.L., Polyakov S.L. Model of Optimal Division of Sides Resources During Information Conflict. *Systems of Control, Communication and Security*. 2018;4:323–344. (in Russ.)
38. Makarenko S.I., Ryimshin K.Yu., Mikhailov R.L. Model of Functioning of Telecommunication Object in the Limited Reliability of Communication Channel Conditions. *Information Systems and Technologies*. 2014;86(6):39–147. (in Russ.)

Сведения об авторах:

МИХАЙЛОВ Роман Леонидович | кандидат технических наук, научно-педагогический сотрудник Военного университета радиоэлектроники, cvviur6@mil.ru

УДК 621.396 DOI:10.31854/1813-324X-2020-6-3-28-37

Алгоритмы выделения канального ресурса в гетерогенной сети радиодоступа нового поколения

С.С. Полевич[1], **О.А. Симонина**[2*]

[1]ООО «Бегет»,
Санкт-Петербург, 195112, Российская Федерация
[2]Санкт-Петербургский государственный университет телекоммуникаций,
Санкт-Петербург, 193232, Российская Федерация
*Адрес для переписки: olga.simonina@spbgut.ru

Информация о статье
Поступила в редакцию 26.08.2020
Принята к публикации 07.09.2020

Ссылка для цитирования: Полевич С.С., Симонина О.А. Алгоритмы выделения канального ресурса в гетерогенной сети радиодоступа нового поколения // Труды учебных заведений связи. 2020. Т. 6. № 3. С. 28–37. DOI:10.31854/1813-324X-2020-6-3-28-37

Аннотация: *В статье предложены алгоритмы выделения канальных ресурсов в гетерогенной сети радиодоступа, использующей одновременно частотные ресурсы сети 5G и сети беспроводного доступа нелицензированного диапазона. Показано, что использование балансировщика в такой сети позволит нивелировать проблему взаимного влияния беспроводных технологий, в частности – интерференции. Для реализации алгоритмов предлагается использовать возможности облачной серверной архитектуры.*

Ключевые слова: *5G, гетерогенная сеть радиодоступа, канальный ресурс, балансировщик.*

Введение

Необходимость модернизации существующих сетей мобильной связи обусловлена ростом трафика, бурным развитием технологий, в том числе автоматизации и удаленного управления объектами, и ростом потребностей пользователя. При этом уже довольно активно ведутся работы по разработке и развертыванию сети сотовой связи нового поколения 5G для обеспечения желаемых показателей – скорости передачи данных, ширине используемого частотного диапазона и многих др.

Одной из важных задач при разработке любой новой технологии является формирование целевой функции сети. Согласно [1], в сетях пятого поколения планируется работа значительного количества устройств, в том числе всевозможных датчиков с высокими требованиями по надежности и пропускной способности. Использование устройств IoT (*от англ.* Internet of Things – Интернета вещей) приведет к еще большему росту трафика, и, как следствие, к более высоким требованиям к сетям нового поколения. В [2] на основании таких прогнозов формируются требования к сетям сотовой связи нового поколения, а также целевые функции сети: повышение пропускной способности; уменьшение задержек (< 1 мс); увеличение количества одновременных подключений устройств к сети оператора; расширение частотного спектра; поддержка большого количества классов качества обслуживания; обеспечение надежной защиты передаваемых данных; постоянная доступность сети; легкость внедрения новых услуг и технологий; снижение количества наращиваемой аппаратной инфраструктуры; эффективное использование ресурсов сети; эффективная доставка контента; снижение энергопотребления абонентскими терминалами; динамическая конфигурируемость сети и т. д.

Операторы сотовой связи мотивируют свою деятельность получением прибыли от предоставления услуг сотовой связи, поэтому можно спрогнозировать ожидаемые экономически выгодные функции сети нового поколения: уменьшение количества узкоспециализированного оборудования; сокращение сигнального трафика и расходов на него; совместимость сети нового поколения с существующей сетью сотовой связи; новые подходы в использовании или переиспользовании частотного спектра; отсутствие необходимости замены оборудования при переходе к 5G и др.

Решения, предлагаемые ведущими вендорами оборудования 5G, строятся на основании облачных решений, охватывающих все уровни сети от ядра до радиодоступа.

Анализ решений организации нижних уровней 5G

В существующих мобильных сетях абонентские терминалы, размещенные на краю соты, имеют беспроводное соединение более низкого качества, основными причинами чего являются удаленность от обслуживающей базовой станции и повышенный уровень интерференции [3]. Внутри помещений, особенно современной застройки, отличающейся большим количеством отражающих поверхностей, надежность покрытия также остается неудовлетворительной. Поэтому более перспективные сети пятого поколения вынуждены применять агрессивные механизмы переиспользования спектрального ресурса и совершенствовать средства управления интерференцией.

Беспроводной спектр является ключевым ресурсом, однако регулярно отмечается [4], что в силу недостаточной зрелости технологий радиодоступа подходы к управлению спектральными ресурсами в последние десятилетия практически не менялись. Развитие разработок в данном направлении приостанавливается существующей необходимостью контролировать множество параметров радиоспектра, таких, как место и время использования, выделенные радиочастоты, излучаемая мощность и возникающая при этом интерференция, которая строго регулируется. При этом многие из этих параметров регламентированы международным законодательством, и внесение изменений в механизмы управления одним из них повлечет за собой как необходимость изменений в нормативно-правовой базе, так и пересмотр норм и требований к другим параметрам.

В настоящее время существуют следующие способы управления спектром:

– командный, представляющий собой статическое выделение спектральных радиоресурсов, а именно, назначение радиочастот с жестко заданными ограничениями на их использование, например, в военных целях; использование спектра при этом неэффективно, так как выделение ресурса не основано на гибких стратегиях переиспользования, механизмах выделения и конкуренции;

– эксклюзивное пользование спектром – это долгосрочная аренда заданной полосы частот для использования в четко оговоренных целях; задается допустимый диапазон частот, область их использования, а также мощность передачи; данный метод позволяет достичь хорошего качества услуг, но в регионах с высокой степенью проникновения радиослужб существенно затрудняет развитие и внедрение новых технологий;

– общее пользование исходно лицензированным спектром подразумевает предоставление диапазонов частот некоторой лицензированной радиослужбы в пользование другим службам без лицензии; при данном способе управления спектром вторичные службы не могут предоставить какие-либо гарантии на качество обслуживания;

– общее пользование нелицензированным спектром представляет собой использование некоторой полосы частот в открытом доступе. Примерами подобных частотных диапазонов являются полосы сети Wi-Fi, когда пользователи получают совместный доступ к беспроводному спектру без дополнительных препятствий; недостатком такого способа управления спектром является неконтролируемая интерференция, существенно понижающая качество услуг.

Для преодоления негативного влияния замедленного развития способов управления спектром требуется разработка новых механизмов организации доступа к радиочастотам общего пользования [5], а также выделение дополнительного радиоресурса для систем мобильной связи. Таким образом, технология 5G должна сочетать в себе целый ряд усовершенствований и новых функций, чтобы повысить производительность беспроводных решений радиодоступа [6]. Все это позволяет наметить основные шаги в направлении развития гетерогенных сетей радиодоступа.

Во-первых, требуется более тесная интеграция различных технологий радиодоступа. Такие решения уже существуют и представляют собой как взаимодействие микро- и макроуровня в сетях мобильной связи, так и агрегацию *WLAN* и сетей мобильной связи [7, 8]. Во-вторых, требуется уплотнение сети, механизмы которого довольно разнообразны и активно совершенствуются из-за необходимости работы в сетях высокой плотности [9]. В-третьих, требуется использовать прямые соединения между устройствами для разгрузки ядра сети, а также высокие частоты для повышения скорости передачи данных (этот подход, к сожалению, имеет очень ограниченную область применения). Также в сети связи нового поколения предполагается использование массивных многоантенных систем *MIMO* и технологии одновременной передачи и приема данных. Среди перечисленных решений уплотнение сети, массивные системы *MIMO* и связь на крайне высоких частотах, то есть миллиметровых волнах (*mmWave*), были названы «большой тройкой» технологий 5G [10].

Применение *mmWave* становится основным нововведением в системах 5G. Это нововведение должно также обеспечить 20-кратное увеличение производительности за счет широких диапазонов малоиспользуемого спектра, доступного в полосе частот от 30 до 300 ГГц и, в меньшей степени, от 20 до 30 ГГц. Однако при внедрении технологий придется преодолеть некоторые сложности, связанные с распространением миллиметровых волн, такие как потери на трассе, низкие коэффициенты дифракции и проникновения, сильный фазовый шум, поглощение атмосферой, ослабление в

дождь, туман или снег [11]. Несмотря на особенности распространения, крайне высокие частоты возможно использовать для систем связи 5G в различных диапазонах [12]. В частности, предполагается использовать следующие частоты: 28–30, 71–76, 81–86 и 92–95 ГГц [13], а также нелицензированные частоты 60 ГГц, уже давно применяющиеся системами связи *IEEE*.

Как отмечалось выше, на данном этапе развития сотовые системы, работающие в лицензированном спектре частот, совмещаются с сетями, работающими в нелицензированном спектре, такими как *Wi-Fi* [14, 15]. В связи с этим современные абонентские терминалы разрабатываются с расчетом возможности использования нескольких технологий радиодоступа [16, 17], что позволяет разгружать сотовые сети посредством прямых соединений между устройствами в нелицензированном спектре [18].

На сегодняшний день многие компании пытаются занять рынок оборудования 5G, несмотря на отсутствие однозначных решений по гармонизации использования спектра. Таким образом, оборудование сетей пятого поколения должно гибко реагировать не только на возможные изменения объема трафика и требований к нему, но и позволять гибко управлять канальным ресурсом как на сети радиодоступа, так и на уровне агрегации и распределения и ядре сети.

Одним из ведущих вендоров в области разработок, адаптированных под сети 5-го поколения, является компания *Huawei* с проектом *Fusion Shere*. *Fusion Shere* – это платформа виртуализации нового поколения, представляющая собой сервер, сетевое оборудование и систему хранения данных в едином корпусе. Согласно анализу [19], современное серверное оборудование не является достаточно мощным и ресурсо-оптимизированным, чтобы стать фундаментом для 5G на основе технологий *NFV* и *SDN*. Эту проблему и решает проект *Huawei*, который отходит от системы кластеров, объединяя серверы в одну целую оболочку. Делается это за счет распределения системы хранения данных. При увеличении количества серверов и хранящейся на них информации производительность сохраняется за счет того, что серверы хранят кэш не в виде физической информации, а в виде хеш-суммы.

По мнению компании *ZTE*, более высокие требования следует предъявлять к транспортной сети. Таким образом, *«Flexhaul» ZTE* – это технология организация транспортной сети, которая должна ввести новые транспортные интерфейсы, технологии и возможности управления сетью, адаптированные к различным сетевым архитектурам, а также удовлетворить требования к пропускной способности, задержкам, высокоточной синхронизации. Идея проекта заключается в поддержке изоляции служб (уровень услуг), а гипервизор распределяет ресурсы между изолированными слоями (уровень логических сетей), но при этом на физическом уровне остаются общие функции управления сетью.

Совместный проект *HP* и *Intel* под названием «*OpenNFV*» представляет собой сервер *Carrier-Grade* и решение *Intel ON Preference Architecture*, построенные на основе технологий *NFV* и *SDN* для предоставления телематических услуг связи. Функциональные блоки системы и интерфейсы взаимодействия между блоками были взяты из наработок Европейского института стандартизации электросвязи (*ETSI, European Telecommunications Standards Institute*) и обеспечивают техническую базу для «*Open NFV*».

Nokia «Cloud Band» – программное обеспечение и архитектура сетевой инфраструктуры, разработанная компанией *NOKIA* на основе решений *NFV* и *SDN*. Система работает с программным обеспечением *Open Stack*, имеющим открытый исходный код. *Open Stack* выполняет роль гипервизора, управляя сетью, распределяя ресурсы и осуществляя контроль за работоспособностью системы в целом. Второй логический элемент системы – это программное обеспечение «*Cloud Band Infrastructure Software*». *Infrastructure Software* управляет виртуальными машинами и их созданием. Установка компонентов, как и всей архитектуры, осуществляется посредством панели инструментов *CBIS Manager* [20].

Таким образом, анализ решений ведущих вендоров показывает, что упор в развитии технологий 5G делается на использование облачных технологий и виртуализации на различных сетевых уровнях. Также имеет смысл отметить активное использование программных решений, обеспечивающих гибкость архитектуры и позволяющих организовать балансировку нагрузки на любом сегменте сети. Результаты анализа характеристик рассмотренных решений приведены в таблице 1.

ТАБЛИЦА 1. Качественные и количественные характеристики решений вендоров

TABLE 1. Qualitative and Quantitative Vendor Solutions Characteristics

Вендор	Критерий			
	Пропускная способность сети, Гбит/с	Задержки	Совместимость с современными решениями	
			программными	аппаратными
Huawei «Fusion Shere»	54	< 1 мс	Да	Нет
ZTE «Flexhaul»	100	< 1 мкс	Нет	Нет
HP and Intel «Open NFV»	10	< 10 мкс	Да	Да
Nokia «Cloud Band»	10	< 10 мкс	Нет	Да

Анализ изменений методов выделения канального ресурса в 5G

Однако гибкое выделение канального ресурса является приоритетной задачей при реализации сценариев обслуживания трафика в сетях пятого поколения. Предполагается, что разнообразные подключенные устройства IoT получат возможность взаимодействовать друг с другом посредством различных технологий радиодоступа. Это обусловлено технологическими решениями и необходимостью удовлетворения целевых функций сети, поэтому понадобится согласованное использование различных радиотехнологий [21], затрагивающее дальнейшее развитие *LTE Advanced* и системы *New Radio (3GPP)* совместно с различными протоколами доступа семейства *IEEE*.

Сеть 5G, как и *LTE*, можно представить в качестве совокупности соединенных между собой станций *gNB* (или *gNodeB* – базовая станция). Радиоинтерфейс между *UE* (*от англ.* User Equipment – абонентское оборудование) и gNB предполагается строить на основе следующих технологий:

– *SCMA* (*от англ*. Sparce Code Multiple Access) – передача битовых потоков всех пользователей в одном частотном ресурсе, которые преобразуются в кодовое слово посредством кодовой книги;

– *F-OFDM* (*от англ.* Flexibel OFDM) – усовершенствованный OFDM, с гибким разбиением на поднесущие, с гибким изменением длины символов и циклического префикса; под каждую задачу используется свой набор параметров.

Таким образом, ядро сети сотовой связи нового поколения представляет собой сервер с виртуальными сетевыми элементами. Коммутация между элементами и со всеми узлами сети предполагается посредством API на основе HTTP, что позволяет заменить множество других протоколов, например, Diametr. Именно на уровне ядра будет обеспечена функция «network slice». Программное обеспечение ядра должно обработать входящий запрос так, чтобы направить его к тем виртуальным элементам сети, которые специализируются на предоставлении соответствующего запросу типа службы/услуг.

При этом к новым сетям относительно *LTE* предъявляются гораздо более жесткие требования к задержкам и пропускной способности.

Для достижения поставленных целей рассматриваются следующие ключевые изменения:

– внедрение и поддержка новых миллиметровых диапазонов, например, согласно *Rel-15 3GPP* будут поддерживаться частоты 37–40 ГГц;

– увеличение полосы пропускания, начиная от 400 МГц и выше;

– варьирование интервала между поднесущими от 15 до 240 кГц в зависимости от частот используемого канала;

– внедрение и поддержка минислотов в кадре: их длительность предполагается меньше чем у слота *LTE*, что позволит уменьшить задержки.

Соответственно, основные методы удовлетворения требованиям 5G – это использование увеличенных полос пропускания базовыми станциями в более широком спектре совместно с новыми технологиями эфирного интерфейса, а также оптимизация задержек за счет внедрения минислотов в кадре и снижения объема передаваемой служебной информации в кадре.

Несмотря на то, что изменения в архитектуре сети приводят к эволюции IP-архитектуры 4G до облачной архитектуры 5G, в сетях нового поколения нет существенных изменений относительно передаваемой информации и методов выделения канального ресурса. Сохранены формы сигналов на основе мультиплексирования с ортогональным частотным разделением, несмотря на внедрение минислотов в кадре длительность кадра сохраняется равной 10 мс, канальный ресурс в 5G делится на ресурсные блоки, каждый из которых состоит из 12 поднесущих. Исходя из этих данных, можно предположить, что распределение символов в кадре также не претерпит изменений, что сделает возможным совместимость вниз с технологиями *LTE* и *LTE Advanced*. Можно сделать вывод о единой структуре методов выделения канального ресурса в сетях нового поколения. На основании этих данных строится новая концепция выделения канального ресурса в 5G.

Алгоритм работы балансировщика канального ресурса в гетерогенной сети радиодоступа

Концепция выделения канального ресурса в 5G заключается в решении задачи оптимального выбора радиотехнологии для предоставления услуги при использовании сотовой и беспроводной локальной сети радиодоступа. Предполагается, что локальная сеть управляется некоторым сотовым оператором, как это происходит в случае использования гетерогенных сетей. Входными данными задачи могут являться следующие показатели: требования к качеству обслуживания, запрашиваемый контент, расположение абонента, интерференция и т. д. На стороне SDN-системы оператора данные поступают в гипервизор, или балансировщик нагрузки, который, приняв решение о распределении ресурсов, рекомендует абонентскому терминалу нужную сеть. В этом же и заключается принцип «*network slice*» в решениях вендоров.

Таким образом, динамика трафика должна отслеживаться на уровне сессии абонентского устройства. Новая сессия, представляющая собой, например, поток данных реального времени с некоторыми минимальными требованиями к скорости передачи, возникает у абонента случайным образом и заканчивается по истечении обслуживания

[22]. Соответственно, число активных сессий изменяется со временем, что можно называть динамикой трафика на уровне сессий. То есть сессия представляет собой поток данных, вызванный установлением удаленного соединения или мультимедиа передачей.

Смоделировать такие сессии возможно посредством пуассоновских процессов. Например, положение абонентских терминалов может определяться некоторым стохастическим потоком точек. Анализ размещения абонентских терминалов и их трафика позволяет оценить мощность принятого сигнала и интерференцию. Таким образом, пространственно-временной подход позволяет управлять трафиком и выбирать предпочтительную радиотехнологию [23–25].

Для пояснения принципа работы системы рассмотрим пример. Возьмем идеализированную макросоту с радиусом R и базовой станцией в ее геометрическом центре. Внутри соты предполагается наличие точек подключения к локальной сети, а также некоторые микросоты. Для наглядности в рассматриваемой системе не предполагается пересечение слоев (network slice). Предположим, что абонентский терминал одного слоя не интерферирует с передачей в других слоях. Интерференция в макросоте при этом не рассматривается. Модель системы представлена на рисунке 1.

Рис. 1. Модель гетерогенной сети радиодоступа

Fig. 1. Heterogeneous Radio Access Network Model

Каждый абонентский терминал в пределах макросоты передает данные по линии вверх, создавая при этом сессию с минимальной требуемой скоростью передачи данных r_0. То есть сеть допускает новую сессию только в том случае, если имеет достаточное количество радиоресурсов для обслуживания. Каждая текущая сессия i занимает r_0/r_i от общего времени работы системы.

Тогда для всех активных сессий выполняется следующее неравенство [26]:

$$\sum_{\text{все сессии}} \left(\frac{r_0}{r_i}\right) \le \delta,$$

где δ – доступные ресурсы.

В целях упрощения системы предположим, что терминал не меняет расположение в пространстве. Тогда коэффициент передачи канала связи между абонентским терминалом и точкой доступа будет зависеть от расстояния между ними. Обозначим коэффициент передачи канала связи для сессии i как:

$$\gamma_{i,j} = \frac{G}{d_{i,j}{}^k},$$

где d_{ij} – расстояние между абонентским терминалом и точкой доступа; k – экспонента распространения; G – константа распространения радиосигнала [26]. При этом k и G зависят от технологии радиодоступа, а сам коэффициент передачи связывает излучаемую мощность со скоростью передачи данных в канале.

Мощность сигнала p_i, излучаемая абонентским терминалом, связана со скоростью передачи данных r_i через обобщенную формулу Шеннона:

$$r_i = B \log\left(1 + A_{p_i}\right),$$

где p_i – мощность сигнала на выходе радиочастотного усилителя; A и B – масштабирующие коэффициенты, зависящие от технологии радиодоступа.

Каждый слой сети принимает на обслуживание сессию в том случае, если для всех сессий слоя выполняются неравенства:

$$r_i \ge r_0, p_i\gamma_{ij} \le N_0, i \ne j,$$

где γ_{ij} – коэффициент передачи радиоканала; p_i – излучаемая мощность радиосигнала.

То есть выражение $p_i\gamma_{ij}$ не должно вызывать интерференцию выше, чем допустимый уровень шума. Следовательно, балансировщик нагрузки проверяет возможность предоставления минимально требуемой скорости передачи данных и уровень интерференции от мобильного терминала.

С учетом перечисленных выше исходных данных устанавливается соединение с сетью по следующему алгоритму:

1) сеть пытается произвести выгрузку новой сессии на обслуживание точки доступа беспроводной локальной сети (*WLAN*), слой *WLAN* при этом может находиться за пределами макросоты, в рамках которой происходит соединение; если сессия попадает в рассматриваемый слой, то она обслуживается локальной сетью на всем жизненном цикле сессии без прерываний, после чего покидает систему;

2) если сессия не может быть принята слоем локальной сети, то сеть пытается выгрузить сессию в микросоту, тогда сессия обрабатывается ближайшей малой БС, а если это невозможно, то сессия не принимается;

3) если локальный и микрослой не способны принять сессию, ее пытается обработать базовая станция макросоты; если сессия не может быть

принята макрослоем, то она окончательно блокируется и покидает систему без обслуживания; на следующие сессии при этом не оказывается никакого влияния.

Графически алгоритм установления соединения представлен на рисунке 2.

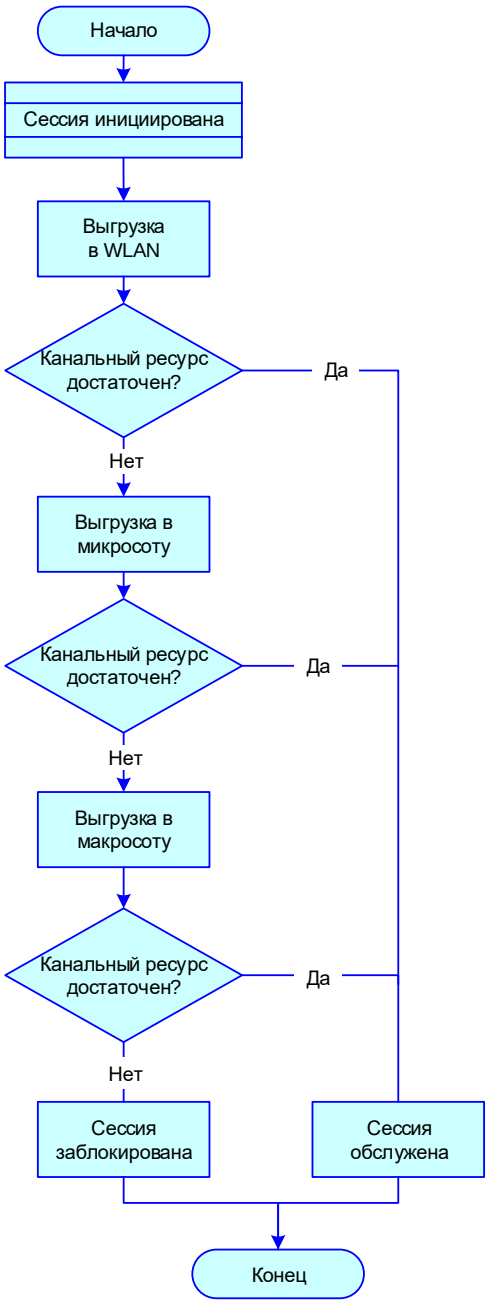


Рис. 2. Алгоритм обслуживания сессии в гетерогенной сети радиодоступа

Fig. 2. The Session Service Algorithm in a Heterogeneous Radio Access Network

Тогда алгоритм балансировки на уровне сессий выглядит следующим образом:

1) определяется мощность передачи абонентского терминала (*UE*), а также возможность предоставления минимальной требуемой скорости $C_{\text{треб}}$ передачи данных при излучаемой мощности;

2) проверяется условие, что ресурс сети делится между абонентскими терминалами поровну;

3) каждая сессий из общего их количества n, допущенная к обслуживанию, получает равную долю от общего ресурса:

$$C = \frac{1}{n}.$$

4) к абонентскому терминалу отправляется сигнал установить мощность передачи $P_{UE\text{треб}}$ для обеспечения минимально требуемой скорости передачи данных.

Блок-схема алгоритма балансировки представлена на рисунке 3.

Ключевым недостатком рассмотренной системы является возможность отказа в обслуживании тех сессий, которые не были допущены в макрослой. Для решения данной проблемы требуется разработка подхода по обслуживанию сессий, которые были заблокированы при первичной попытке доступа в сеть.

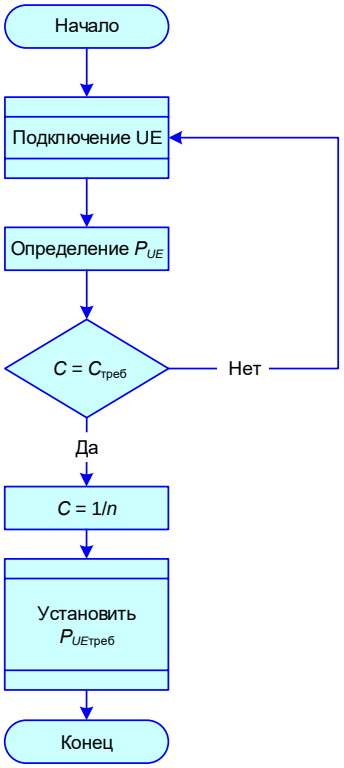


Рис. 3. Блок-схема алгоритма балансировки

Fig. 3. The Block Diagram of the Balancing Algorithm

Алгоритм выделения канального ресурса в гетерогенной сети радиодоступа высокой плотности

Чтобы снизить порог отказа в обслуживании заблокированных сессий, можно уплотнить гетерогенную радиосеть, а также использовать совместно лицензионный и нелицензионный спектр. То есть возможна интеграция, например, технологии *Wi-Fi* как части инфраструктуры сотовой сети, в которой абонентский терминал может передавать данные

одновременно и через лицензионный и нелицензионный спектр [26].

Предлагается плотное размещение малых сот, в которых каждая базовая станция будет оснащена, например, функционалом как мобильной сети, так и *WLAN*. При этом *WLAN* функционирует в нелицензионном диапазоне частот, а макросота служит не для обработки трафика, а для управления системой малых сот и радиоресурсами. Основной особенностью описанной системы является передача данных абонентскими терминалами по линии вверх как по сети мобильной связи, так и параллельно по *WLAN*. В зависимости от требуемого качества обслуживания, абонентский терминал в такой системе может использовать как только одну из технологий, так и одновременно задействовать ресурсы доступных радиотехнологий в возможных пропорциях. В таком случае необходим функциональный блок на стороне сети оператора, который будет принимать решение о распределении ресурсов технологий радиодоступа на основе целевой функции.

Функционал такого блока может быть реализован на основе серверной архитектуры и входить в ядро сети 5G. При этом необходимо ввести ряд проверок, позволяющих корректировать подключение с целью уменьшения интерференции в соседних сотах.

Во-первых, перед тем как принять сессию какой-либо базовой станцией, на стороне управления нагрузкой должна выполниться проверка достижения максимального количества уже обслуживаемых сессий, так как при любой организации сети такой максимум будет существовать.

Во-вторых, после того как сессия была принята, она получает равную долю радиоресурса стандарта мобильной связи, а при необходимости и *WLAN* для организации требуемого качества обслуживания. Сессия, получившая ресурс и требуемую скорость передачи данных, обслуживается без прерываний, после чего немедленно выводится из системы.

Предполагается, что абонентский терминал передает данные на максимальной мощности, если при этом не превышается допустимый уровень интерференции. Если сессия на предоставленных радиоресурсах с заданной мощностью вызывает прирост показателя интерференции в соседней соте, то соответствующие радиоресурсы становятся недоступны для оказания услуг в соседней соте. Основная идея балансировщика нагрузки в такой системе состоит в том, чтобы забрать дополнительные ресурсы для предоставления минимально требуемых для новой или более приоритетной сессии.

Алгоритм работы такой гетерогенной системы представлен на рисунке 4.

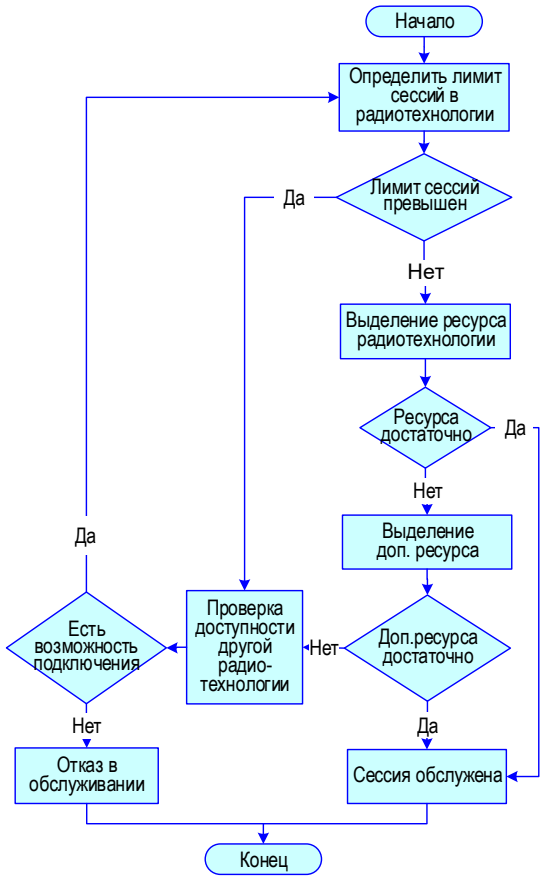


Рис. 4. Алгоритм выделения ресурсов в гетерогенной сети радиодоступа

Fig. 4. The Resource Allocation Algorithm in a Heterogeneous Radio Access Network

Можно предположить, что абонентский терминал будет постоянно обмениваться данными с макросотой и поддерживать соединение для передачи данных с малой сотой. Тогда система связи организует централизованное управление переключением терминала между малыми сотами, что повысит производительность за счет общесистемной информации о состоянии всей сети. Пример реализации балансировщика гетерогенной сети радиодоступа с использованием серверной архитектуры на базе решений с открытой лицензией приведен на рисунке 5.

Выводы

Таким образом, в статье показано, что внесение функционального блока, позволяющего контролировать несколько радиотехнологий доступа на уровне ядра, возможно реализовать с использованием серверной архитектуры. Такой подход позволит организовать гетерогенную сеть 5G, нивелируя проблемы взаимного влияния технологий радиодоступа и одновременно оптимизировать выделение канального ресурса в соответсвтии с требованиями приложений и услуг. Предложенные алгоритмы могут лечь в основу открытого проекта реализации современной гетерогенной сети радиодоступа.

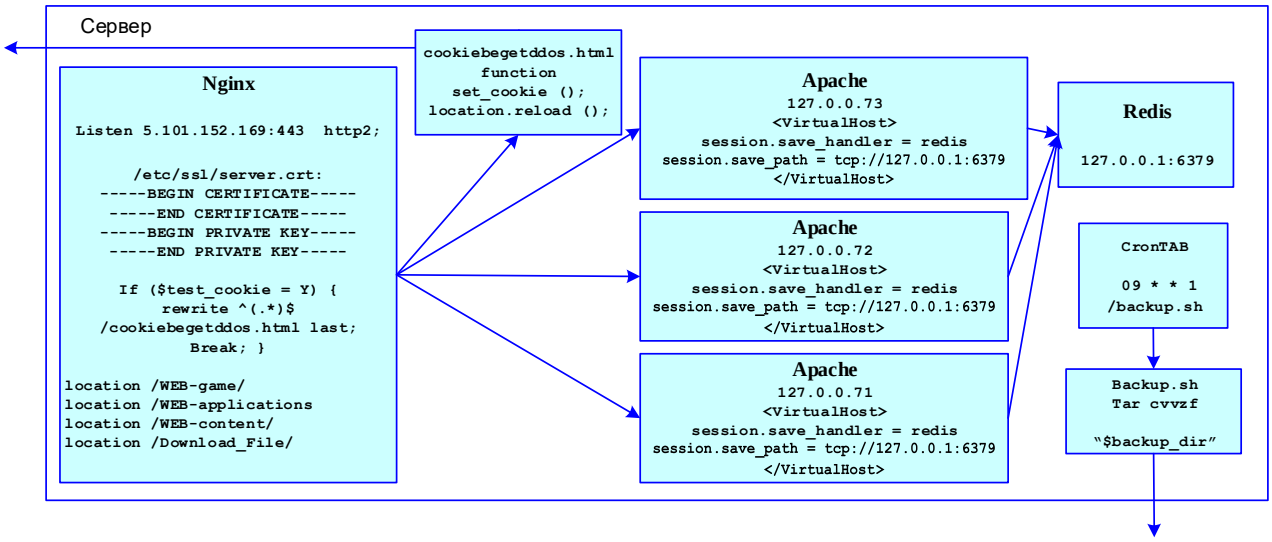


Рис. 5. Возможная реализация балансировщика нагрузки в гетерогенной сети 5G с использованием серверной архитектуры

Fig. 5. Possible Load Balancer Implementation in a Heterogeneous 5G Network Using a Server Architecture

В качестве направлений дальнейших исследований можно отметить потребность в адекватном моделировании сетей связи 5G для обслуживания перспективных сценариев Интернета мобильных вещей с высокими требованиями по надежности подключения. Соответственно, актуальны исследование и разработка гетерогенных радиосистем с возможностью одновременного подключения. В таких системах устройства могут применять как сотовые, так и прямые соединения. В частности, в публикациях наблюдается дефицит результатов моделирования радиосистемы в целом для типовых сценариев 5G-IoT, предполагающих наличие нескольких радиотехнологий доступа, которые пользовательские устройства могут использовать альтернативно или одновременно, и при этом учитывать возможные влияния интерференции между технологиями.

Список используемых источников

1. Mobile World Congress 2019 // Cisco. URL: https://blogs.cisco.com/tag/mwc-2019 (дата обращения 29.04.2020)
2. 3GPP Release 15. Release description. Technical report: 21.915 SA#83 V:1.1.0, 21.06.2019
3. Tombaz S., Vastberg A., Zander J. Energy- and cost-efficient ultra-high-capacity wireless access // IEEE Wireless Communications. 2011. Vol. 18. Iss. 5. PP. 18–24. DOI:10.1109/MWC.2011.6056688
4. Ponomarenko-Timofeev A., Pyattaev A., Andreev S., Koucheryavy Y., Mueck M., Karls I. Highly dynamic spectrum management within licensed shared access regulatory framework // IEEE Communications Magazine. 2016. Vol. 54. Iss. 3. PP. 100–109. DOI:10.1109/MCOM.2016.7432155
5. Mokrov E., Ponomarenko-Timofeev A., Gudkova I., Masek P., Hosek J., et al. Modeling Transmit Power Reduction for a Typical Cell With Licensed Shared Access Capabilities // Transactions on Vehicular Technology. 2018. Vol. 67. Iss. 6. PP. 5505–5509. DOI:10.1109/TVT.2018.2799141
6. Vannithamby R., Talwar S. Towards 5G: Applications, Requirements and Candidate Technologies. Chichester: John Wiley & Sons, 2017. 466 p.
7. Loginov V.A., Lyakhov A.I., Khorov E.M. Coexistence of Wi-Fi and LTE-LAA Networks: Open Issues // Journal of Communications Technology and Electronics. 2018. Vol. 63. PP. 1530–1537. DOI:10.1134/S1064226918120148
8. Wang L., Kuo G.-S.G.S. Mathematical Modeling for Network Selection in Heterogeneous Wireless Networks – A tutorial // IEEE Communications Surveys & Tutorials. 2012. Vol. 15. Iss. 1. PP. 271–292. DOI:10.1109/SURV.2012.010912.00044
9. Ле Ч.Д., Симонина О.А. Механизм мультиопроса на основе приоритизации для WLAN с высокой плотностью устройств // Труды учебных заведений связи. 2017. Т. 3. №. 1. С. 80–92.
10. Andrews J.G., Buzzi S., Choi W., Hanly S.V., Lozano A., Soong A.C.K., et al. What will 5G be? // IEEE Journal on Selected Areas in Communications. 2014. Vol. 32. Iss. 6. PP. 1065–1082. DOI:10.1109/JSAC.2014.2328098
11. Karjalainen J., Nekovee M., Benn H., Kim W., Park J.H., Sungsoo H. Challenges and opportunities of mm-wave communication in 5G networks // Proceedings of the 9th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM, Oulu, Finland, 2–4 June 2014). IEEE, 2014. PP. 372–376. DOI:10.4108/icst.crowncom.2014.255604
12. Rappaport T.S., Sun S., Mayzus R., Zhao H., Azar Y., Wang K., et al. Millimeter Wave Mobile Communications for 5G Cellular: It Will Work! // IEEE Access. 2013. Vol. 1. PP. 335–349. DOI:10.1109/ACCESS.2013.2260813
13. Boccardi F., Heath R.W., Lozano A., Marzetta T.L., Popovski P. Five disruptive technology directions for 5G // IEEE Communications Magazine. 2014. Vol. 52. Iss. 2. PP. 74–80. DOI:10.1109/MCOM.2014.6736746
14. Song W., Zhuang W., Cheng Y. Load balancing for cellular/WLAN integrated networks // IEEE Network. 2007. Vol. 21. Iss. 1. PP. 27–33. DOI:10.1109/MNET.2007.314535

15. Bellalta B., Bononi L., Bruno R., Kassler A. Next generation IEEE 802.11 Wireless Local Area Networks: Current status, future directions and open challenges // Computer Communications. 2016. Vol. 75. PP. 1–25. DOI:10.1016/j.comcom.2015.10.007
16. Zhou Z., Jia Y., Chen F., Tsang K., Liu G., Han Z. Unlicensed Spectrum Sharing: From Coexistence to Convergence // IEEE Wireless Communications. 2017. Vol. 24. Iss. 5. PP. 94–101. DOI:10.1109/MWC.2017.1700086
17. Zetterman T., Piipponen A., Raiskila K., Slotte S. Multi-Radio coexistence and collaboration on an SDR platform // Analog Integrated Circuits and Signal Processing. 2011. Vol. 69. DOI:10.1007/s10470-011-9713-7
18. Sankaran C. B. Data offloading techniques in 3GPP Rel-10 networks: A tutorial // IEEE Communications Magazine. 2012. Vol. 50. Iss. 6. PP. 46–53. DOI:10.1109/MCOM.2012.6211485
19. FusionSphere Virtualization Suite 6.5.0 Product Documentation // Huawei. URL: https://support.huawei.com/hedex/hdx.do?docid=EDOC1100071114&lang=en&idPath=22658044%7C7919788%7C9856606%7C21462752%7C9823560 (дата обращения 30.04.2020)
20. Cloud Band Release 17.5 // NOKIA. URL: https://onestore.nokia.com/asset/200060 (дата обращения 06.05.2020)
21. Gerasimenko M., Moltchanov D., Andreev S., Koucheryavy Y., Himayat N., Yeh S.-P., et al. Adaptive Resource Management Strategy in Practical Multi-Radio Heterogeneous Networks // IEEE Access. 2016. Vol. 5. PP. 219–235. DOI:10.1109/ACCESS.2016.2638022
22. Kim Y., de Veciana G. Joint Network Capacity Region for Cognitive Networks Heterogeneous Environments and RF-Environment Awareness // IEEE Journal on Selected Areas in Communications. 2011. Vol. 29. Iss. 2. PP. 407–420. DOI:10.1109/JSAC.2011.110213
23. Kelly F., Yudovina E. Stochastic Networks. Cambridge: Cambridge University Press, 2014. 232 p.
24. Андреев С.Д., Кучерявый Е.А., Самуйлов К.Е. Пространственно-временной подход к анализу гетерогенных систем связи // Электросвязь. 2018. № 9. С. 20–26.
25. 3GPP Release 8 Technical report: 37.83, 2019.
26. Orsino A., Ometov A., Fodor G., Moltchanov D., Militano L., Andreev S., et al. Effects of Heterogeneous Mobility on D2D- and Drone-Assisted Mission-Critical Mtc in 5G // IEEE Communications Magazine. 2017. Vol. 55. Iss. 2. PP. 79–87. DOI:10.1109/MCOM.2017.1600443CM

* * *

Algorithms for Channel Resource Allocation in a Heterogeneous New-Generation Radio Access Network

S. Polevich[1], **O. Simonina**[2]

[1]Beget Ltd,
St. Petersburg, 195027, Russian Federation
[2]The Bonch-Bruevich Saint-Petersburg State University of Telecommunications,
St. Petersburg, 193232, Russian Federation

Article info
DOI:10.31854/1813-324X-2020-6-3-28-37
Received 26th August 2020
Accepted 7th September 2020

For citation: Polevich S., Simonina O. Algorithms for Channel Resource Allocation in a Heterogeneous New-Generation Radio Access Network *Proc. of Telecom. Universities.* 2020;6(3):28–37. (in Russ.) DOI:10.31854/1813-324X-2020-6-3-28-37

Abstract: *The article proposes algorithms for allocating channel resources in a heterogeneous radio access network that simultaneously uses the frequency resources of a 5G network and an unlicensed wireless access network. It is shown that the use of a balancer in such network will make it possible to level the problem of the mutual wireless technologies influence, in particular, interference. To implement the algorithms, it is proposed to use the capabilities of the cloud server architecture.*

Keywords: *5G, heterogeneous radio access network, channel resource, load balancer.*

References

1. Cisco. *Mobile World Congress 2019*. Available from: https://blogs.cisco.com/tag/mwc-2019 [Accessed 29th April 2020]
2. 3GPP Release 15. *Release description*. Technical report: 21.915 SA#83 V:1.1.0, 21.06.2019

3. Tombaz S., Vastberg A., Zander J. Energy- and cost-efficient ultra-high-capacity wireless access. IEEE Wireless Communications. 2011;18(5):18–24. DOI:10.1109/MWC.2011.6056688
4. Ponomarenko-Timofeev A., Pyattaev A., Andreev S., Koucheryavy Y., Mueck M., Karls I. Highly dynamic spectrum management within licensed shared access regulatory framework // IEEE Communications Magazine. 2016. Vol. 54. Iss. 3. PP. 100–109. DOI:10.1109/MCOM.2016.7432155
5. Mokrov E., Ponomarenko-Timofeev A., Gudkova I., Masek P., Hosek J., et al. Modeling Transmit Power Reduction for a Typical Cell With Licensed Shared Access Capabilities. T*ransactions on Vehicular Technology*. 2018;67(6):5505–5509. DOI:10.1109/TVT.2018.2799141
6. Vannithamby R., Talwar S. *Towards 5G: Applications, Requirements and Candidate Technologies*. Chichester: John Wiley & Sons; 2017. 466 p.
7. Loginov V.A., Lyakhov A.I., Khorov E.M. Coexistence of Wi-Fi and LTE-LAA Networks: Open Issues. *Journal of Communications Technology and Electronics*. 2018;63:1530–1537. DOI:10.1134/S1064226918120148
8. Wang L., Kuo G.-S.G.S. Mathematical Modeling for Network Selection in Heterogeneous Wireless Networks – A tutorial. *IEEE Communications Surveys & Tutorials.* 2012;15(1):271–292. DOI:10.1109/SURV.2012.010912.00044
9. Le Tran Duc, Simonina O. The Multipolling Mechanism Based on the Prioritization for WLAN in a High Dense Networks. *Proc. of Telecom. University*. 2017;3(1):80–92. (in Russ.)
10. Andrews J.G., Buzzi S., Choi W., Hanly S.V., Lozano A., Soong A.C.K., et al. What will 5G be? *IEEE Journal on Selected Areas in Communications*. 2014;32(6):1065–1082. DOI:10.1109/JSAC.2014.2328098
11. Karjalainen J., Nekovee M., Benn H., Kim W., Park J.H., Sungsoo H. *Challenges and opportunities of mm-wave communication in 5G networks. Proceedings of the 9th International Conference on Cognitive Radio Oriented Wireless Networks and Communications, CROWNCOM, 2–4 June 2014, Oulu, Finland*. IEEE; 2014. p.372–376. DOI:10.4108/icst.crowncom.2014.255604
12. Rappaport T.S., Sun S., Mayzus R., Zhao H., Azar Y., Wang K., et al. Millimeter Wave Mobile Communications for 5G Cellular: It Will Work! *IEEE Access*. 2013;1:335–349. DOI:10.1109/ACCESS.2013.2260813
13. Boccardi F., Heath R.W., Lozano A., Marzetta T.L., Popovski P. Five disruptive technology directions for 5G. *IEEE Communications Magazine*. 2014;52(2):74–80. DOI:10.1109/MCOM.2014.6736746
14. Song W., Zhuang W., Cheng Y. Load balancing for cellular/WLAN integrated networks. *IEEE Network*. 2007;21(1): 27–33. DOI:10.1109/MNET.2007.314535
15. Bellalta B., Bononi L., Bruno R., Kassler A. Next generation IEEE 802.11 Wireless Local Area Networks: Current status, future directions and open challenges. *Computer Communications*. 2016;75:1–25. DOI:10.1016/j.comcom.2015.10.007
16. Zhou Z., Jia Y., Chen F., Tsang K., Liu G., Han Z. Unlicensed Spectrum Sharing: From Coexistence to Convergence. *IEEE Wireless Communications*. 2017;24(5):94–101. DOI:10.1109/MWC.2017.1700086
17. Zetterman T., Piipponen A., Raiskila K., Slotte S. Multi-Radio coexistence and collaboration on an SDR platform. *Analog Integrated Circuits and Signal Processing*. 2011;69. DOI:10.1007/s10470-011-9713-7
18. Sankaran C. B. Data offloading techniques in 3GPP Rel-10 networks: A tutorial. *IEEE Communications Magazine*. 2012; 50(6):46–53. DOI:10.1109/MCOM.2012.6211485
19. Huawei. *FusionSphere Virtualization Suite 6.5.0 Product Documentation* Available from: https://support.huawei.com/hedex/hdx.do?docid=EDOC1100071114&lang=en&idPath=22658044%7C7919788%7C9856606%7C21462752%7C9823560 [Accessed 30th April 2020]
20. NOKIA. *Cloud Band Release 17.5*. Available from: https://onestore.nokia.com/asset/200060 [Accessed 6th May 2020]
21. Gerasimenko M., Moltchanov D., Andreev S., Koucheryavy Y., Himayat N., Yeh S.-P., et al. Adaptive Resource Management Strategy in Practical Multi-Radio Heterogeneous Networks. *IEEE Access*. 2016;5:219–235. DOI:10.1109/ACCESS.2016.2638022
22. Kim Y., de Veciana G. Joint Network Capacity Region for Cognitive Networks Heterogeneous Environments and RF-Environment Awareness. *IEEE Journal on Selected Areas in Communications*. 2011;29(2):407–420. DOI:10.1109/JSAC.2011.110213
23. Kelly F., Yudovina E. *Stochastic Networks*. Cambridge: Cambridge University Press; 2014. 232 p.
24. Andreev S.D., Koucheryavy Y.A., Samouylov K.E. Space-Time Analysis of Heterogeneous Wireless Systems. *Elektrosvyaz*. 2018;9:20–26. (in Russ.)
25. 3GPP Release 8. Technical report: 37.83, 2019.
26. Orsino A., Ometov A., Fodor G., Moltchanov D., Militano L., Andreev S., et al. Effects of Heterogeneous Mobility on D2D- and Drone-Assisted Mission-Critical Mtc in 5G. *IEEE Communications Magazine*. 2017;55(2):79–87. DOI:10.1109/MCOM.2017.1600443CM

Сведения об авторах:

ПОЛЕВИЧ Сергей Сергеевич | системный администратор компании «ООО Бегет», sergeypolevich@mail.ru

СИМОНИНА Ольга Александровна | кандидат технических наук, доцент кафедры радиосвязи и вещания Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича, olga.simonina@spbgut.ru

УДК 004.057.4 DOI:10.31854/1813-324X-2020-6-3-38-46

Разработка протокола взаимодействия меток и считывателей для системы позиционирования на основе измерения времени распространения радиосигнала

А.В. Тарлыков[1]

[1]Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича, Санкт-Петербург, 193232, Российская Федерация
*Адрес для переписки: atarlykov@gmail.com

Информация о статье
Поступила в редакцию 10.07.2020
Принята к публикации 11.08.2020

Ссылка для цитирования: Тарлыков А.В. Разработка протокола взаимодействия меток и считывателей для системы позиционирования на основе измерения времени распространения радиосигнала // Труды учебных заведений связи. 2020. Т. 6. № 3. С. 38–46. DOI:10.31854/1813-324X-2020-6-3-38-46

Аннотация: *В статье рассматривается задача построения протокола взаимодействия подвижных меток и стационарных считывателей для системы позиционирования, основанной на измерении времени распространения радиосигнала. Проанализирован вопрос уменьшения количества передаваемых сообщений. Приводится общее описание предлагаемого протокола, структура сообщений и основные сценарии взаимодействия.*

Ключевые слова: *позиционирование, метки, протокол, время распространения радиосигнала.*

Введение

Системы позиционирования все сильнее проникают в повседневные области деятельности. Это могут быть задачи, связанные с контролем переносного оборудования, оптимизацией деятельности персонала, контролем доступа в опасные и не предусмотренные регламентом зоны, навигация в закрытых пространствах и другие задачи, более комплексные и базирующиеся на выше перечисленных. Для решения подобных задач могут использоваться различные технологии, к примеру, достаточно распространенным подходом является использование RFID-меток [1]. Но подобные метки не рассчитаны на отслеживание точного местоположения объектов и, как правило, применяются для общего контроля, на их основе могут строиться системы учета времени и общий мониторинг. Другим подходом является использование систем RTLS (*от англ.* Real Time Locating System) [2, 3]. Главное их достоинство – отслеживание координат контролируемого объекта с достаточно высокой (до сантиметров) точностью и достоверностью контроля позволяют RTLS-системам решать гораздо более широкий круг задач в сравнении с другими.

Общая схема рассматриваемых RTLS-систем

Основной задачей рассматриваемой системы позиционирования является контроль и отслеживание местоположения подвижных объектов. Каждый отслеживаемый объект ассоциируется с аппаратной меткой, местоположение которой отслеживается относительно стационарных считывателей. Объем контролируемой зоны определяется геометрическим расположением считывателей. Таким образом, в состав системы позиционирования входят следующие основные элементы (рисунок 1): стационарные считыватели с фиксированными координатами; подвижные метки; центральная система расчета положения меток на основе координат считывателей и расстояний до части из них.

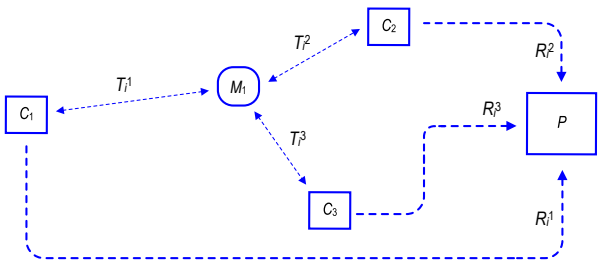


Рис. 1. Общая схема взаимодействия элементов системы

Fig. 1. Interactions of System Elements

На рисунке 1 приведены элементы системы, участвующие в отдельно взятой транзакции по определению местоположения метки, где: C_j – считыватели; M_i – метка, инициировавшая транзакцию; T_i^j – измеренное в процессе транзакции время распространения сигнала от метки M_i до считывателя C_j; R_i^j – рассчитанное считывателем C_j расстояние до метки M_i; P – централизованная система расчета положения меток.

Считыватели и метки характеризуются уникальными идентификаторами, задаваемыми на этапе начальной настройки системы. Метки являются подвижными, основную часть времени находятся в состоянии сна, пробуждаются через определенные промежутки времени и инициируют транзакцию с рядом считывателей для определения расстояний. Набор и количество сообщений в рамках каждой транзакции может варьироваться и определяется протоколом. По окончании транзакции все считыватели, успешно рассчитавшие расстояние до метки, передают информацию в центральную систему для расчета обновленного положения метки. В случае возникновения коллизий транзакция завершается по тайм-ауту.

В процессе работы метка может взаимодействовать с любыми доступными считывателями. В рамках же отдельно взятой транзакции у метки существует возможность исключить ряд считывателей из взаимодействия, что, в ряде случаев, должно позволить разгрузить радиоканал и повысить точность за счет использования различных наборов опорных считывателей.

Необходимо отметить, что в основе функционирования подобных систем лежит механизм доступа к среде ALOHA и условием стабильной работы является минимизация утилизации радиоканала, что, в свою очередь, ведет к необходимости уменьшения количества передаваемых сообщений и их размера. Данную задачу и пытается решить предлагаемый в статье протокол.

Методы позиционирования в RTLS-системах

В рамках RTLS-систем используются различные методы позиционирования, отличающиеся наборами измеряемых параметров. Среди них можно выделить три, используемых наиболее часто [2–6]:

– по уровню принимаемого радиосигнала, RSSI (*от англ.* Received Signal Strength Indicator) [7, 8];

– по углам относительно известных направлений [9];

– на основе измерения времени распространения радиосигнала.

Технологии позиционирования на основе измерения времени распространения радиосигнала часто объединяются под общим названием ToF (*от англ.* Time of Flight) [5]. В рамках данного метода существует целый ряд технологий определения расстояний, различаются они способом получения и обработки данных, на основе которых производится расчет местоположения радиометки. В основном метод базируется на технологиях беспроводной связи, подпадающих под действие группы стандартов IEEE 802.15.4a [10, 11], поддерживающих важную опцию измерения времени распространения сигнала между приемником и передатчиком. Так как измерение времени производится одновременно с пересылкой данных, то поддержка данной функции не требует дополнительных затрат энергии и большей пропускной способности каналов. Кроме того, появляется возможность одновременного решения задач позиционирования и параметрического мониторинга. Приведем основные технологии функционирования подобных систем:

– технология TDoA (*от англ.* Time Difference of Arrival): определение расстояния производится по разнице во времени распространения радиосигнала от радиометки до группы считывателей, что требует синхронизации часов приемной части системы; радиометка с определенной частотой отсылает сообщения на все «видимые» считыватели, которые определяют моменты времени прихода радиосигнала и передают эту информацию на сервер обработки, где и производятся вычисления местоположения метки как пересечение трех и более гипербол;

– технология ToA (*от англ.* Time of Arrival): в данном случае используется тактовая синхронизация всей приемо-передающей аппаратуры, точность синхронизации достаточно важна и может доходить до нескольких пикосекунд; радиометка передает радиосигнал в фиксированные моменты времени; считыватели измеряют время между отправкой радиосигнала радиометкой и его приемом считывателем, далее рассчитывается расстояние; высокие требования к синхронизации приводят к сложностям при эксплуатации и росту стоимости решений;

– технология TWR (*от англ.* Two Way Ranging [12]). В данном случае используется два сообщения: первое – от метки к считывателю, и второе – от считывателя к метке. Время задержки ответа на считывателе фиксируется и отсылается в ответном сообщении метке, таким образом, метка рассчитывает расстояние на основе зафиксированных моментов передачи и приема сообщений, а также – времени задержки на считывателе. Тактовая синхронизация при данной технологии не используется, так как отсылка и прием радиосигнала осуществляются одним и тем же устройством. Данная технология не лишена проблем и основная из них – нестабильность опорных генераторов, участвующих в обмене устройств, что приводит к трудностям измерения положения с точностью более метра [13];

– технология DS-TWR (*от англ.* Double Sided – Two Way Ranging [12]): в схеме DS-TWR процесс измерения времени производится дважды (рисунок 2): радиометка инициирует процесс измерения и замеряет время T_1 между отсылкой и приемом ответного сообщения, затем аналогичную операцию выполняет считыватель, замеряя время T_3. Далее расстояние рассчитывается на основе времен $T_1 - T_4$, где T_2 и T_4 – времена задержки ответов считывателем и радиометкой, соответственно; как и в варианте TWR, времена T_2 и T_4 пересылаются в ответных сообщениях и доступны для использования в расчетах;

– технология SDS-TWR (от англ. Symmetrical Double Sided – Two Way Ranging [3, 12]): разновидность DS-TWR с равными (фиксированными) временами ответа обоих устройств (T_2 и T_4 на рисунке 2), что упрощает вычисления, но привносит ряд трудностей – поочередный обмен метки со всеми считывателями и увеличение времени итерации для поддержки максимального расстояния.

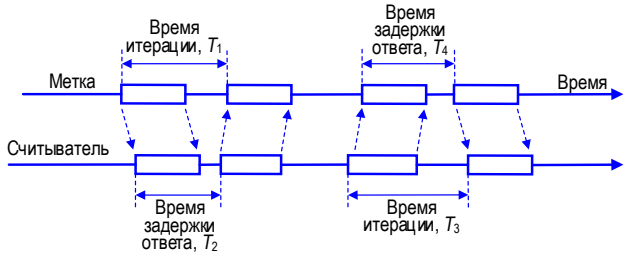


Рис. 2. Пример обмена в режиме DS-TWR
Fig. 2. Example of Interactions in DS-TWR Mode

Таким образом, вариантов взаимодействия группы меток и считывателей существует достаточно много. Но необходимо отметить, что приведенные выше варианты рассматривают взаимодействие одной метки и небольшой группы считывателей, достаточной для определения местоположения метки. Непосредственное использование указанных схем является далеким от оптимальности решением для практических сценариев, включающих большое количество меток и считывателей.

Далее, при построении протокола, в качестве базиса будет использоваться технология DS-TWR [14]. Кроме базового варианта с использованием четырех сообщений для определения расстояния, данная технология позволяет использовать несколько оптимизаций. Во-первых, существует возможность использовать не четыре сообщения для определения расстояния метка – считыватель, а три (рисунок 3), что повышает общую эффективность схемы. В данном случае ответное сообщение от считывателя используется для измерения времени сразу двух итераций обмена: метка – считыватель – метка и считыватель – метка – считыватель.

Более того, в данном случае нет необходимости включать время T_2 (задержка ответа считывателем) в ответное сообщение метке. Все расчеты могут производиться считывателем, который запоминает время T_2 на своей стороне и получает в ответном сообщении от метки времена T_1 и T_4. В данном случае автоматически решается и проблема передачи от метки рассчитанного на ее стороне расстояния для первой итерации метка – считыватель – метка. Все расчеты выполняет считыватель, что также снижает и суммарный объем передаваемых через радиоканал данных.

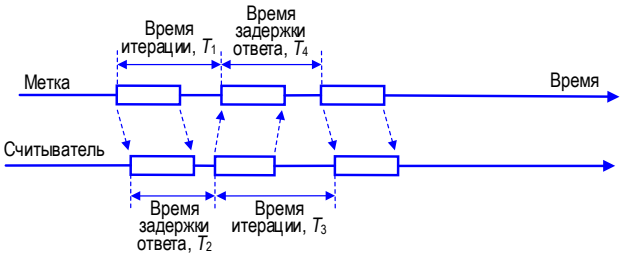


Рис. 3 Пример обмена с использованием трех сообщений
Fig. 3. Interactions with Three Messages

Второй важной особенностью режима DS-TWR является возможность группировки сообщений, что связано с отсутствием требования синхронности режима SDS-TWR (времена T_2 и T_4 не фиксированы, см. рисунки 2, 3), что, в свою очередь, позволяет в схеме с одной меткой и тремя считывателями уменьшить общее количество сообщений с девяти до пяти – одно инициирующее сообщение ко всем считывателям, три ответных от считывателей и единый ответ метки всем считывателям (рисунок 4).

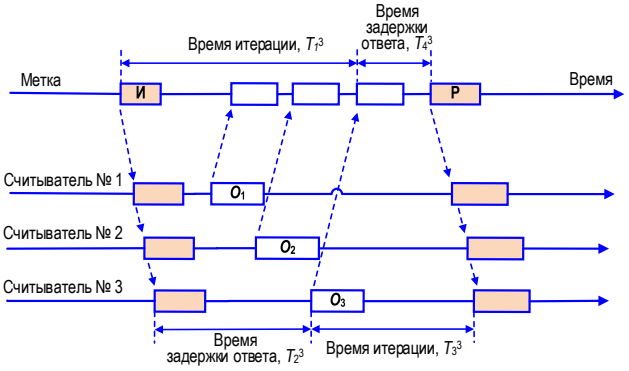


Рис. 4. Пример обмена с использованием пяти сообщений
Fig. 4. Interactions with Five Messages

Здесь запись вида T_4^i определяет время T_4 (в соответствии с рисунками 2, 3) при взаимодействии метки со считывателем с индексом i.

Протокол взаимодействия меток и считывателей

Рассмотрим вариант построения протокола обмена между совокупностью меток и считывателей на основе технологии DS-TWR с использованием группировки сообщений. Целью будет являться уменьшение общего количества сообщений, необходимых для измерения положения всех меток.

Все сообщения будут содержать в себе стандартный набор полей и иметь следующий вид:

$$|FC, APP, FUNC, FDATA|,$$

где поля определяются следующим образом:

– FC (Frame Control), управляющая информация для сообщения, может задавать битовую размерность адресов устройств, модификаторы протокола и т. п.;

– APP, уникальный идентификатор системы или приложения;

– $FUNC$, код прикладного сообщения;

– $FDATA$, данные текущего сообщения, формат определяется значением поля $FUNC$.

В дальнейших рассуждениях поля FC и APP будут опускаться при описании частных сообщений протокола как не относящиеся к логике взаимодействия меток и считывателей.

При реализации протокола будем ссылаться на ряд параметров, значения которых могут варьироваться при необходимости (таблица 1).

ТАБЛИЦА 1. Временные параметры протокола

TABLE 1. Temporal Parameters of the Protocol

Параметр	Описание
T_{tp}	Период запросов со стороны метки, инициирующих процесс измерения расстояния до ряда считывателей, может быть как фиксированным, так и варьироваться в некоторых пределах
T_{tps}	Период запросов со стороны метки, инициирующих процесс измерения расстояния для уточнения положения в случае необходимости, $T_{tps} \le T_{tp}$
T_{tcl}	Длительность временного окна ожидания меткой ответов от считывателей после посылки инициирующего сообщения об измерении положения, вариант повышенной длительности (см. далее)
T_{tc}	Длительность временного окна ожидания меткой ответов от считывателей, вариант обычной длительности (см. далее)
T_{apl}	Длительность временного окна для посылки ответа считывателем после приема инициирующего сообщения от метки (вариант повышенной длительности), $T_{apl} \le T_{tcl}$
T_{ap}	Длительность временного окна для посылки ответа считывателем после приема инициирующего сообщения от метки (вариант обычной длительности), $T_{ap} \le T_{tc}$

Далее рассмотрим общую логику поведения метки и типы сообщений, необходимые для решения задачи позиционирования. В процессе работы метка периодически инициирует транзакции по измерению расстояния до ряда считывателей. В простейшем варианте это могут быть все установленные считыватели или считыватели, принявшие инициирующее сообщение. Каждая транзакция включает в себя инициирующее сообщение метки, ответы от считывателей и результирующее сообщение от метки (технология DS-TWR позволяет осуществлять группировку ответных сообщений). Очевидно, присутствует линейная зависимость общего количества сообщений в рамках транзакции от количества ответивших считывателей, и важным является минимизация подобных сообщений в процессе работы. Варианты минимизации в большей степени определяются логикой построения алгоритма считывателя, но и на уровне протокола должна существовать возможность подобную логику реализовать.

Будем использовать несколько вариантов сообщения для инициирования транзакции. Самым простым вариантом выступает широковещательное сообщение о начале транзакции, участвовать в которой может любой считыватель, принявший запрос. Данный вариант может быть использован в качестве «холодного старта» при включении метки или метка может «откатиться» к его использованию при падении количества отвечающих считывателей в процессе обмена. Транзакция, инициируемая данным типом сообщения, характеризуется использованием временного окна повышенной длительности (см. таблицу 1). Формат сообщения следующий:

$$|FM1, TID, SEQ|,$$

где $FM1$ – код сообщения инициации транзакции; TID – идентификатор метки; SEQ – последовательный циклический номер транзакции метки.

Второй вариант сообщения инициирующего транзакцию выглядит полностью аналогично и отличается только кодом сообщения ($FM2$). Тем не менее, данный вариант является основным и позволяет считывателям использовать различные механизмы для минимизации количества ответов. Также данный тип транзакции характеризуется использованием временного окна уменьшенной длительности для накопления ответов от считывателей.

При использовании сообщений типа $FM2$ возможен вариант, когда по какой-либо причине метке ответило мало считывателей или неудачно расположенные считыватели. Например, в случае, если метка получила ответ от двух считывателей, чего недостаточно для полноценного позиционирования. В подобных случаях следующая транзакция может быть организована меткой ранее очередного запланированного сеанса.

Также разумной выглядит возможность явного исключения ряда считывателей из участия в очередной транзакции, что может быть использовано как для уточнения положения при недостаточном количестве участников в предыдущей транзакции, так и для разгрузки среды передачи, исключая ответы ряда считывателей. Для поддержки подобной возможности предлагается расширить набор инициирующих сообщений метки дополнительным вариантом, включающим идентификаторы соответствующих считывателей:

$$|FM3, TID, SEQ, \#A, AID\{\#A\}|,$$

где $FM3$ – код сообщения инициации транзакции с исключением ряда считывателей; $\#A$ – количество считывателей, исключаемых из инициируемого сеанса измерения расстояния; AID – идентификаторы считывателей в количестве, соответствующем значению поля $\#A$.

Подобное расширение набора инициирующих сообщений позволяет явно менять наборы считывателей в последовательных транзакциях (для уточнения положения относительно разных наборов считывателей), что может быть актуально при плотном покрытии территории считывателями и при наличии различных помех в направлении ряда считывателей. Но важно отметить, что подобная поддержка может выходить за возможности конкретной метки, например, метка может быть оборудована малым объемом памяти, не позволяющим вести статистику использования считывателей.

Стоит отдельно упомянуть возможный вариант инициирующего сообщения метки с указанием конкретных целевых считывателей (обратный вариант к $FM3$). На первый взгляд такой вариант должен выступать простым способом кардинального снижения количества ответных сообщений от считывателей. Но при дальнейшем рассмотрении оказывается, что на метку переходит задача по выбору предпочтительных считывателей, что приводит к повышению нагрузки на метку и росту требований к аппаратному обеспечению, в первую очередь – объему памяти. Также метка ограничена в возможностях построения оптимального набора считывателей для обмена в рамках транзакции, сказывается весьма ограниченный объем доступной ей информации. Возможным вариантом могло бы являться использование набора считывателей из предыдущей транзакции в течение нескольких итераций, что требует минимальных затрат от метки. Но отсутствие гарантий на оптимальность набора считывателей, участвовавших в любой отдельно взятой транзакции, сильно снижает эффективность подобного подхода. Таким образом, возможность всех разом расположенных считывателей на каждом итерации «соревноваться» за участие в транзакции позволяет повысить вероятность построения «удачных» комбинаций считывателей в ходе последовательных транзакций. Также отказ от указания целевых считывателей в сообщении позволяет переложить задачу оптимизации на часть системы (считыватели), обладающую априори большими ресурсами и большей информацией о состоянии всей системы. Все перечисленные выше моменты указывают на нецелесообразность использования варианта инициирующего сообщения с фиксированным набором целевых считывателей.

Следующая стадия обработки транзакции приходится на считыватели. Каждый из них большую часть времени находится в режиме приема сообщений. После приема, инициирующего сообщения от какой-либо метки, считыватель случайным образом выбирает время ответа в пределах длительности окна ответа (соответствующего типу принятого сообщения) и помечает данное событие в своем плане ответов. Наличие плана ответов позволяет реализовать параллельный обмен с несколькими метками и, более того, производить упаковку сообщений в ряде случаев. В случае, когда помещаемый в план ответ пересекается по времени передачи с уже запланированным ответом другой метке, такие ответы могут быть объединены в единое сообщение для нескольких меток. Подобное решение позволяет уменьшить суммарное время передачи и пересечения с другими транзакциями.

Таким образом, ответные сообщения считывателя в рамках транзакций измерения расстояния с метками могут принимать один из двух вариантов. Первый вариант нацелен на общение с одной меткой и выглядит следующим образом:

$$|FA1, AID, TID, SEQ|$$

где $FA1$ – код ответного сообщения для одной метки; AID – идентификатор считывателя; TID – идентификатор метки, которой предназначен ответ; SEQ – текущий номер транзакции, полученный от метки.

Второй вариант реализует возможность ответа нескольким меткам:

$$|FA2, AID, \#T, \ (TID^i, SEQ^i)\{\#T\}|,$$

где $FA2$ – код ответного сообщения для группы меток; $\#T$ – количество целевых меток; (TID^i, SEQ^i) – наборы из номера транзакции и идентификатора соответствующей метки в количестве, указанном в поле $\#T$.

Рассмотрим дальнейшее функционирование метки. После отсылки инициирующего сообщения какого-либо вида метка переходит в режим приема для обработки ответов от считывателей. На прием отводится окно времени фиксированной длительности (см. таблицу 1). Если инициирующее сообщение позволяло оптимизацию ответов, то в ходе приема ответных сообщений метка может закрыть окно приема после получения определенного количества ответов (например, трех ответов уже может быть достаточно для определения местоположения). В случае отсутствия ответных сообщений в течение открытого окна приема, метка завершает транзакцию и переходит в режим сна до следующей попытки. Кроме ответов на инициирующее сообщение самой метки, метка может принять ответы считывателей другим меткам. В общем случае подобные сообщения могут быть проигнорированы, но в ряде случаев могут быть использованы для определения расстояния в самом простом режиме TWR с использованием всего двух сообщений и

увеличенной погрешности. Данный вариант может быть актуален в случае недостаточного количества ответов самой метке.

Использование ассиметричной схемы DS-TWR позволяет метке «упаковать» ответ всем считывателям в единое сообщение. Закрыв окно приема (заранее или по истечении времени), метка отсылает ответ всем считывателям, сообщения от которых были получены. Общая структура сообщения выглядит следующим образом:

$$|FM3, TID, SEQ, \#AR, (AID^i, T_1^i, T_4^i)\{\#AR\}, \#AS, (AID^i, T_4^i)\{\#AS\}|,$$

где $FM3$ – код сообщения окончания транзакции; $\#AR$ – количество считывателей, сообщения от которых были получены в рамках транзакции и которым предназначено данное финальное сообщение; $\#AS$ – количество считывателей, сообщения которых были предназначены другим меткам, были «подслушаны» данной меткой и с которыми метка предлагает измерить расстояние в режиме обмена двумя сообщениями (данный блок может отсутствовать); (AID^i, T_1^i, T_4^i) – наборы, включающие идентификатор считывателя, время обмена сообщениями метка – считыватель – метка и время задержки между приемом ответного сообщения соответствующего считывателя и общим ответом (текущее сообщение), индексы соответствуют рисунку 3, количество наборов соответствует значению поля $\#AR$; (AID^i, T_4^i) – наборы из идентификатора считывателя и времени задержки между приемом ответного сообщения соответствующего считывателя и общим ответом (текущее сообщение), индексы соответствуют рисунку 3, количество наборов соответствует значению поля $\#AS$.

После приема финального сообщения от метки, считывателем производится расчет расстояния и отсылка информации в отслеживающую систему по одному из доступных каналов связи. Возможны следующие состояния транзакции с меткой в момент приема финального сообщения от нее:

– считыватель находится в списке адресатов в принятом сообщении и существует активная транзакция с данной меткой: такое состояние означает, что от метки был принят запрос на измерение расстояния и ей был отослан ответ (индивидуальный или в составе группового ответа) в рамках стандартного процесса обмена;

– считыватель находится в списке адресатов в принятом сообщении, активная транзакция с меткой отсутствует: в данном случае от метки не было запросов на измерение расстояния (или он не был принят), но она приняла ответ данного считывателя некой другой метке и решила использовать его для оценки расстояния в режиме двух сообщений (вариант TWR). Данное решение метки может быть продиктовано, к примеру, недостаточным количеством ответов от считывателей на ее начальный запрос измерения расстояний;

– считыватель не находится в списке адресатов сообщения, но транзакция с меткой активна и существует запланированный ответ в плане ответов: данная ситуация может сложиться в случае, когда ответ считывателя запланирован ближе к окончанию временного окна ответа метке, а метка накопила необходимое (или максимальное) количество ответов от других считывателей и приняла решение о досрочном закрытии окна приема ответных сообщений от считывателей. В данном случае считыватель удаляет запланированный ответ метке из плана ответов, что пересекается с общей логикой функционирования считывателя в случае мониторинга сообщений от других считывателей (как было указано выше), в данном случае ответ метке может быть отменен заранее;

– также возможен вариант утери финального сообщения в результате коллизий в радиоканале, в подобном случае транзакция завершается считывателем по истечении времени: при необходимости информация может передаваться в центральную систему для мониторинга количества сбоев и потерь в работе системы.

Как упоминалось выше, основной вклад в минимизацию количества передаваемых сообщений дает уменьшение ответов считывателей на инициирующие запросы меток. Количество сообщений остальных типов уменьшению практически не подлежат, возможна только оптимизация путем их агрегации, поддержка чего была предложена в базовом протоколе.

Рассматривая оптимизацию количества ответов со стороны считывателей, можно предложить следующие варианты. Считыватель должен осуществлять контроль ответных сообщений других считывателей в рамках каждой активной транзакции измерения расстояния и отменять отсылку своего ответа в случае, когда к моменту ответа накоплено определенное количество ответов данной метке от других считывателей (конкретное значение является параметром протокола). Данный вариант оптимизации позволяет отсечь ответные сообщения, размещенные ближе к окончанию временного окна ответа метке (в плане ответов) и согласуется с предварительным закрытием данного окна со стороны метки.

Следующим шагом оптимизации может стать вариант с периодическим пропуском части транзакций на основе информации о расстоянии до метки, предполагая, что с увеличением расстояния возрастает вероятность нахождения метки в окрестности других считывателей. Также возможен контроль служебных сообщений других считывателей (например, с информацией о рассчитанном расстоянии до меток) и на основании этих данных принятие решения об ответе метке на

инициирующее сообщение. Возможен и вариант обмена сообщениями между считывателями, что может быть использовано для взаимной оценки расстояния и использования этих данных для автоматического построения схемы взаимного расположения считывателей. Подобная схема позволяет отслеживать примерное положение меток в окрестности считывателя и принимать решение об участии в транзакции в зависимости от расположения каждой метки относительно совокупности считывателей.

Часть указанных вариантов оптимизации реализуется достаточно просто и полностью основываются на базовом протоколе взаимодействия считывателей и меток. Оставшаяся часть, которая требует использования дополнительного протокола для взаимодействия самих считывателей, реализуется сложнее, но также не требует доработки базового протокола.

Сравним количество сообщений, необходимое для определения положения каждой метки для различных вариантов взаимодействия меток и считывателей (таблица 2). Будем считать достаточным измерение расстояния до трех считывателей для удачного определения местоположения каждой метки. Обозначим количество меток через T и количество считывателей – через A.

Могут возникнуть опасения, что группировка приводит к сильному увеличению размера сообщений и повышает вероятность коллизий. Детальный анализ загруженности радиоканала и вероятности коллизий выходит за рамки данной статьи, но приведем оценку времени занятости радиоканала для случаев индивидуальных и группового ответа метки четырем считывателям. Обратимся к рекомендациям «ГОСТ Р 58082–2018» [15] на системы позиционирования в реальном времени. Рассмотрим предлагаемые в разделе № 5.2 режимы функционирования метки с длиной преамбулы 256 и 1024 символов при скорости передачи данных – 850 кб/с. Дополнительно проведем анализ для скорости передачи 6,81 Мбит/с, входящей в стандарт и поддерживаемой, например, модулями Decawave DW1000 [16]. Будем считать размер блока данных равным 16 байтам для индивидуальных ответов метки одному считывателю и 56 байтам для группового ответа одновременно четырем считывателям. Используя данные из раздела «5.3.4 Временные параметры преамбулы» [15] примерно оценим время, необходимое для пересылки данных считывателям (таблица 3).

На приведенном примере можно видеть, что группировка сообщений существенно сокращает суммарное время передачи и, соответственно, загрузку радиоканала. Основной выигрыш достигается за счет экономии на передаче таких служебных полей сообщений, как преамбулы и блоки синхронизации. Аналогичным образом выигрыш достигается и для остальных типов сообщений с группировкой передачи. Также поддержка протоколом исключения набора считывателей из транзакции с меткой позволяет снизить конкуренцию считывателей за радиоканал при ответе метке и снижает общее количество коллизий.

ТАБЛИЦА 2. Количество сообщений для различных вариантов взаимодействия

TABLE 2. Number of Messages for Various Modes

Вариант	Сообщений, шт.	Сообщений (T = 100, A = 20), шт.
Режим четырех сообщений каждой пары метка – считыватель	$4AT$	8 000
Режим четырех сообщений, но единым инициирующим сообщением	$T * (1 + 3A)$	6 100
Режим четырех сообщений, но единым инициирующим сообщением и группировкой ответов меткой	$T * (4 + 2A)$	4 400
Режим трех сообщений и единым инициирующим сообщением	$T * (4 + A)$	2 400
Режим трех сообщений, единым инициирующим сообщение и группировкой ответов меткой	$T * (2 + A)$	2 200
Рассмотренный протокол без оптимизации со стороны считывателей	$T * (2 + A)$	2 200
Рассмотренный протокол без оптимизации со стороны считывателей и перекрытием 10 % запросов от меток	$T * (2 + 0{,}9A)$	2 000
Рассмотренный протокол с функцией мониторинга со стороны считывателей	$5T$	500
Рассмотренный протокол с функцией мониторинга со стороны считывателей и перекрытием 10 % запросов от меток	$4{,}5T$	450

ТАБЛИЦА 3. Оценка времени передачи данных считывателям

TABLE 3. Summarized TX Time to Notify Four Anchors

Размер преамбулы, символов	Скорость передачи блока данных	Суммарное время индивидуальных ответов, мкс	Время группового ответа, мкс
256	850 Кбит/с	1 732	773
	6,81 Мбит/с	1 100	349
1024	850 Кбит/с	4 788	1 537
	6,81 Мбит/с	4 276	1 113

В рамках дальнейших исследований планируется провести полноценное моделирование системы позиционирования с использованием предлагаемого протокола для более точной оценки загрузки радиоканала и возможных коллизий.

Таким образом, использование протокола и достаточно простых оптимизаций на его основе позволяет принципиально снизить количество необходимых сообщений в процессе определения местоположения меток и снизить суммарное время занятости радиоканала. С увеличением количества участников – меток и считывателей – эффект увеличивается, что должно снизить количество коллизий передаваемых сообщений по сравнению с простыми вариантами взаимодействия. Что, в общем итоге, должно позволить создавать более сложные и объемные комплексы для измерения местоположения объектов.

Список используемых источников

1. Want R. An introduction to RFID technology // IEEE Pervasive Computing. 2006. Vol. 5. Iss. 1. PP. 25–33. DOI:10.1109/MPRV.2006.2
2. Gaffney B. Considerations and Challenges in Real Time Locating Systems Design // Decawave. URL: https://www.decawave.com/considerations-and-challenges-in-real-time-locating-systems-design (дата обращения 03.03.2020)
3. Real Time Location Systems (RTLS) // Nanotron Technologies GmbH. URL: https://nanotron.com/assets/pdf/whitepapers/WP_RTLS.pdf (дата обращения 03.03.2020)
4. Кривченко Т. Программно-аппаратные методы измерения расстояния по времени распространения радиосигнала при помощи приемопередатчика nanoLOC // Беспроводные технологии. 2012. № 3(28). С. 48–53.
5. Botta M., Simek M., Krajsa O., Cervenka V., Pal T. On Location Estimation Technique Based of the Time of Flight in Low-power Wireless Systems // Measurement Science Review. 2015. Vol. 15. Iss. 2. PP. 58–63. DOI:10.1515/msr-2015-0009
6. Миниахметов Р.Х., Рогов А.А., Цымблер М.Л. Обзор алгоритмов локального позиционирования для мобильных устройств // Вестник Южно-Уральского государственного университета. Серия: Вычислительная математика и информатика. 2013. Т. 2. № 2. С. 83–96.
7. Sauter M. From GSM to LTE-Advanced: An Introduction to Mobile Networks and Mobile Broadband: Second Edition. 2014. DOI:10.1002/9781118861943
8. iBeacon // Apple Developer. URL: https://developer.apple.com/ibeacon (дата обращения 03.03.2020)
9. Dave H. How AoA & AoD Changed the Direction of Bluetooth Location Services // Bluetooth. URL: https://www.bluetooth.com/blog/new-aoa-aod-bluetooth-capabilities (дата обращения 03.03.2020)
10. Karapistoli E., Pavlidou F.-N., Gragopoulos I., Tsetsinas I. An overview of the IEEE 802.15.4a Standard // Communications Magazine. 2010. Vol. 48. Iss. 1. PP. 47–53. DOI:10.1109/MCOM.2010.5394030
11. IEEE Std 802.15.4-2011. IEEE Standard for Local and metropolitan area networks. Part 15.4: Low-Rate Wireless Personal Area Networks (LR-WPANs). DOI:10.1109/IEEESTD.2011.6012487
12. Sang C.L., Adams M., Hörmann T., Hesse M., Porrmann M., Rückert U. Numerical and Experimental Evaluation of Error Estimation for Two-Way Ranging Methods // Sensors. 2019. Vol. 19. Iss. 3. DOI:10.3390/s19030616
13. Decawave. Sources of error in DW1000 based two-way ranging (twr) schemes // APS011 application note. URL: https://www.decawave.com/wp-content/uploads/2018/08/aps011_sources_of_error_in_twr.pdf (дата обращения 03.03.2020)
14. Decawave. The implementation of two-way ranging with the dw1000 // APS013 application note. URL: https://www.decawave.com/wp-content/uploads/2018/08/aps013_dw1000_and_two_way_ranging_v2.2.pdf (дата обращения 03.03.2020)
15. ГОСТ Р 58082–2018 (ИСО/МЭК 24730-62: 2013). Системы позиционирования в реальном времени (RTLS). Часть 62. Сверхширокополосный радиоинтерфейс с высокой частотой повторения импульсов. М.: Стандартинформ, 2018. 64 с.
16. Decawave. DW1000 User Manual. How to use, configure and program the dw1000 uwb transceiver. URL: https://www.decawave.com/sites/default/files/resources/dw1000_user_manual_2.11.pdf (дата обращения 03.03.2020)

* * *

Developing Anchor-Tag Communication Protocol for Positioning System Based on Time of Flight Measurement

A. Tarlykov[1]

[1]The Bonch-Bruevich Saint-Petersburg State University of Telecommunications,
St. Petersburg, 193232, Russian Federation

Article info
DOI:10.31854/1813-324X-2020-6-3-38-46
Received 10th July 2020
Accepted 11th August 2020

For citation: Tarlykov A. Developing Anchor-Tag Communication Protocol for Positioning System Based on Time of Flight Measurement. *Proc. of Telecom. Universities.* 2020;6(3):38–46. (in Russ.) DOI:10.31854/1813-324X-2020-6-3-38-46

Abstract: *The article deals with the problem of designing a protocol for the interaction of movable tags and stationary readers based on measuring the spreading time of a radio signal. The issue of reducing the number of transmitted messages is analyzed. Common description of the proposed protocol, the messages and main interaction scenarios structure are provided.*

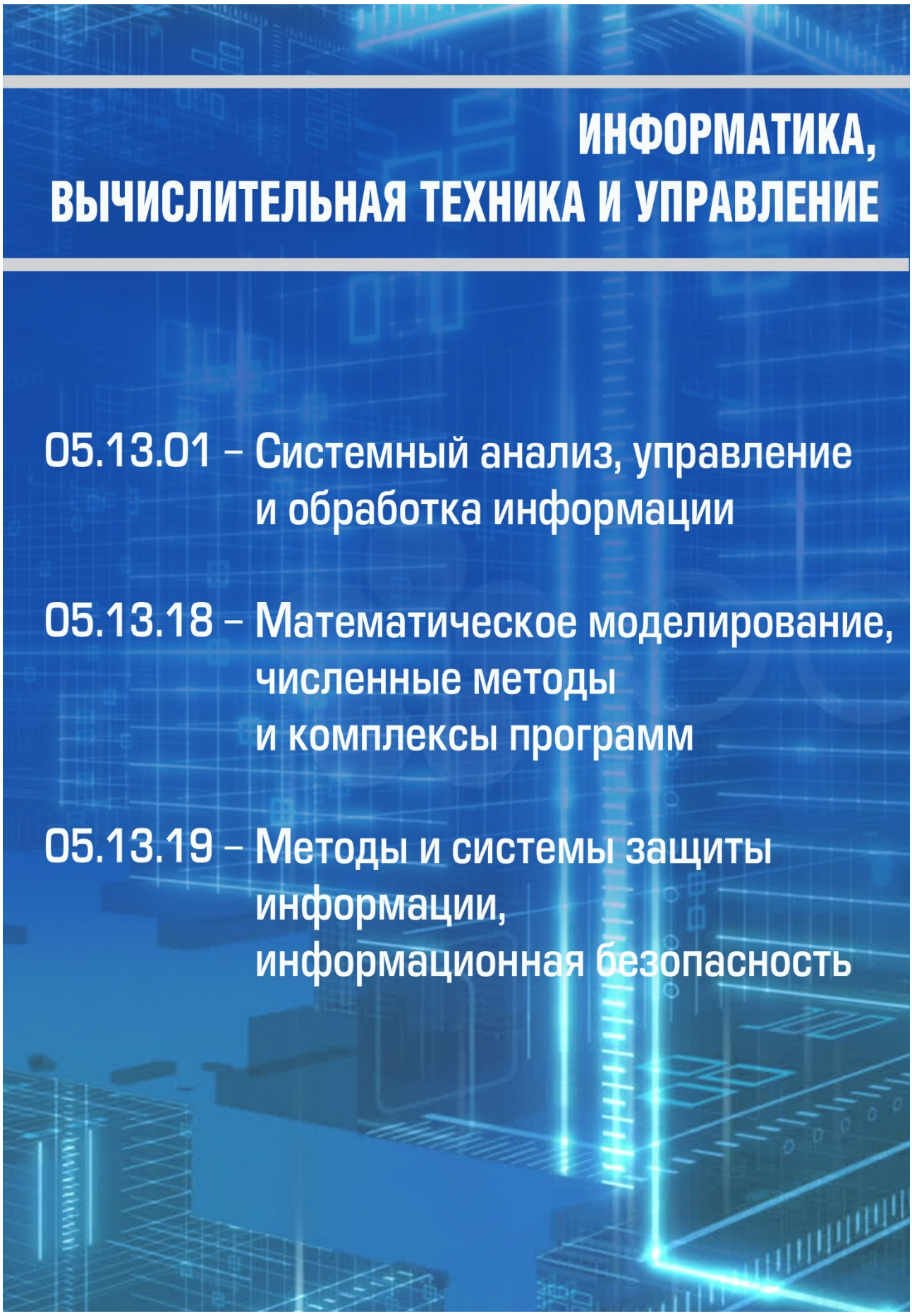
Keywords: *positioning, tags, anchors, protocol, time of flight.*

References

1. Want R. An introduction to RFID technology. *IEEE Pervasive Computing*. 2006;5(1):25–33. DOI:10.1109/MPRV.2006.2
2. *Decawave*. Gaffney B. Considerations and Challenges in Real Time Locating Systems Design. Available from: https://www.decawave.com/considerations-and-challenges-in-real-time-locating-systems-design [Accessed 3rd March 2020]
3. Real Time Location Systems (RTLS). *Nanotron Technologies GmbH*. Available from: https://nanotron.com/assets/pdf/whitepapers/WP_RTLS.pdf [Accessed 3rd March 2020]
4. Krivchenko T. Software and Hardware Methods for Measuring the Distance in Terms of the Propagation Time of a Radio Signal Using the nanoLOC Transceiver. *Wireless technologies*. 2012;3(28):48–53. (in Russ.)
5. Botta M., Simek M., Krajsa O., Cervenka V., Pal T. On Location Estimation Technique Based of the Time of Flight in Low-power Wireless Systems. *Measurement Science Review*. 2015;15(2):58–63. DOI:10.1515/msr-2015-0009
6. Miniakhmetov R.M., Rogov A.A., Zymbler M.L. The Survey of Indoor Positioning Algorithms for Mobile Devices. Bulletin of the South Ural State University. Series "Computational Mathematics and Software Engineering". 2013;2(2):83–96. (in Russ.)
7. Sauter M. *From GSM to LTE-Advanced: An Introduction to Mobile Networks and Mobile Broadband: Second Edition*. 2014. DOI:10.1002/9781118861943
8. Apple Developer. *iBeacon*. Available from: https://developer.apple.com/ibeacon [Accessed 3rd March 2020]
9. *Bluetooth*. Dave H. How AoA & AoD Changed the Direction of Bluetooth Location Services. Available from: https://www.bluetooth.com/blog/new-aoa-aod-bluetooth-capabilities [Accessed 3rd March 2020]
10. Karapistoli E., Pavlidou F.-N., Gragopoulos I., Tsetsinas I. An overview of the IEEE 802.15.4a Standard. *Communications Magazine*. 2010;48(1):47–53. DOI:10.1109/MCOM.2010.5394030
11. *IEEE Std 802.15.4-2011*. IEEE Standard for Local and metropolitan area networks. Part 15.4: Low-Rate Wireless Personal Area Networks (LR-WPANs). DOI:10.1109/IEEESTD.2011.6012487
12. Sang C.L., Adams M., Hörmann T., Hesse M., Porrmann M., Rückert U. Numerical and Experimental Evaluation of Error Estimation for Two-Way Ranging Methods. *Sensors*. 2019;19(3). DOI:10.3390/s19030616
13. Decawave. Sources of error in DW1000 based two-way ranging (twr) schemes. *APS011 application note*. Available from: https://www.decawave.com/wp-content/uploads/2018/08/aps011_sources_of_error_in_twr.pdf [Accessed 3rd March 2020]
14. Decawave. The implementation of two-way ranging with the dw1000. *APS013 application note*. Available from: https://www.decawave.com/wp-content/uploads/2018/08/aps013_dw1000_and_two_way_ranging_v2.2.pdf [Accessed 3rd March 2020]
15. GOST P 58082–2018 *Information technology. Real time locating systems (RTLS). Part 62. High rate pulse repetition frequency Ultra Wide Band (UWB) air interface*. Moscow: Standartinform Publ.; 2020. (in Russ.)
16. Decawave. DW1000 User Manual. How to use, configure and program the dw1000 uwb transceiver. URL: https://www.decawave.com/sites/default/files/resources/dw1000_user_manual_2.11.pdf [Accessed 3rd March 2020]

Сведения об авторе:

ТАРЛЫКОВ Алексей Владимирович | начальник НОЦ «Лаборатория программирования» Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича, atarlykov@gmail.com
https://orcid.org/0000-0002-5707-2406



DOI:10.31854/1813-324X-2020-6-3

УДК 004.056.5 DOI:10.31854/1813-324X-2020-6-3-48-57

Идентификация архитектуры процессора выполняемого кода на базе машинного обучения. Часть 3. Оценка качества и границы применимости

М.В. Буйневич[1, 2], **К.Е. Израилов**[1, 3*]

[1]Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича, Санкт-Петербург, 193232, Российская Федерация
[2]Санкт-Петербургский университет государственной противопожарной службы МЧС России, Санкт-Петербург, 196105, Российская Федерация
[3]Санкт-Петербургский федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация
[*]Адрес для переписки: konstantin.izrailov@mail.ru

Информация о статье
Поступила в редакцию 10.07.2020
Принята к публикации 08.09.2020

Ссылка для цитирования: Буйневич М.В., Израилов К.Е. Идентификация архитектуры процессора выполняемого кода на базе машинного обучения. Часть 3. Оценка качества и границы применимости // Труды учебных заведений связи. 2020. Т. 6. № 3. С. 48–57. DOI:10.31854/1813-324X-2020-6-3-48-57

Аннотация: *В статье изложены результаты тестирования авторского способа идентификации архитектуры процессора выполняемого кода на базе машинного обучения. В третьей заключительной части цикла определены его качественные показатели: точность, полнота и F-мера для выполняемых файлов сборки Debian. Исследованы границы применимости способа идентификации архитектуры для четырех условий: отсутствие заголовка файлов, различные размеры машинного кода, частичное разрушение кода и наличие инструкций нескольких архитектур. Указаны обнаруженные недостатки предлагаемого способа и пути их устранения, а также дальнейшее направление развития.*

Ключевые слова: *информационная безопасность, машинный код, архитектура процессора, машинное обучение, сигнатура кода, способ идентификации, показатели качества, матрица ошибок, точность, полнота, F-мера, разрушение кода.*

Введение

Одной из задач информационной безопасности является исследования файлов информационной системы, наиболее важными из которых являются выполняемые – содержащие машинный код (далее – МК) [1]. Первоначальным шагом такого исследования служит определение архитектуры процессора МК, что даст общую картину о составе информационной системы, а также позволит выбрать наиболее подходящий инструментарий для ее анализа [2–4]; притом даже в случае разрушения файловой структуры [5]. Существуют различные способы решения данной задачи, рассмотренные в 1-ой части цикла статей [6]. Недостатки каждого из них привели к необходимости создания частотно-байтовой модели кода, а затем и к разработке авторского способа идентификации на базе машинного обучения, описанного во 2-ой части цикла [7].

Продолжая следовать канонической схеме «анализ → синтез → оценка», в данной статье будет произведена оценка качества [8] способа идентификации, а также определены границы его применимости. В интересах первого действия будут получены точность, полнота и *F-мера* способа, а в интересах второго – исследованы работоспособность способа при отсутствии заголовка файлов [9, 10], зависимость от размера кода и степени его разрушения [11], а также влияние нескольких процессорных архитектур в коде. Затем потребуется обсуждение полученных результатов, из чего последуют предложения касательно дальнейшего усовершенствования способа.

Оценка качества классификации

Показатели качества

Наиболее часто используемыми показателями качества классификации (применимыми к множеству задач, например, сетевой безопасности [12]

или обработки текста [13]) являются *точность* и *полнота*. Применительно к задаче определения архитектуры процессора МК, первая определяет долю файлов, для которых архитектура была определена верно, по сравнению со всеми файлами, отнесенными способом к данной архитектуре. Вторая определяет долю файлов с верно определенной архитектурой по сравнению со всеми истинными файлами данной архитектуры. Естественно, идеальным случаем будет максимальное значение точности и полноты, что в реальных условиях практически недостижимо; по этой причине, как правило, вводится *F-мера*, представляющая собой некий компромисс между точностью и полнотой в виде среднего гармонического. Более же полную картину о проведенной идентификации архитектур для некоторой тестовой выборки можно увидеть по *матрице ошибок*, отражающей в численном виде взаимосвязь между реальными архитектурами процессора и теми, которые определил способ. Все остальные показатели качества могут быть напрямую получены из этой матрицы, поэтому целесообразно получить вначале именно ее.

Матрица ошибок

Для получения матрицы ошибок авторское программное средство идентификации архитектуры процессора МК было модернизировано добавлением 4-го режима работы – «Вычисление метрик», в котором, помимо сигнатур и архитектур процессора МК, средству указывались директории с выполняемыми файлами и реальными архитектурами процессоров. Таким образом, производится сравнение архитектур, полученных с помощью разработанного способа, и реальных.

Так, после запуска программного средства в данном режиме (1-ый аргумент *MetricMode*) с помощью командной строки:

```
> BinArchId.exe MetricMode 3 amd64=D:\Debian\Amd64\ i386=D:\Debian\I386\
mipsel=D:\Debian\Mipsel\ amd64=D:\Debian\amd64.sig i386=D:\Debian\i386.sig
                          mipsel=D:\Debian\mipsel.sig
```

оно выполнит 5 действий. Во-первых, установит число исследуемых архитектур, равное 3 (2-ой аргумент). Во-вторых, вычислит сигнатуры файлов из директорий, вложенных в D:\Debian\, указав для них верные архитектуры – *amd64*, *i386* и *mipsel* (с 3-го по 5-ый аргументы). В-третьих, аналогично режиму «Идентификация МК» (при аргументе *TestMode*), загрузит из директории «D:\Debian\» сигнатуры «*amd64.sig*», «*i386.sig*» и «*mipsel.sig*» для соответствующих архитектур (аргументы с 6-го по 8-ой). В-четвертых, произведет идентификацию всех файлов согласно сигнатурам. И, в-пятых, построит матрицу ошибок, где столбцами является реальная архитектура процессора МК, строками – идентифицированная архитектура, а значениями в ячейках – количество файлов с обеими архитектурами. В случае первого аргумента *MetricModeX*, файлы будут интерпретироваться как полностью состоящие из МК (т. е. без учета заголовка и секций). Для получения более корректных оценок режим использует одинаковое количество файлов каждой из архитектур – путем вычисления их минимального значения из всех.

Для вычисления матрицы ошибок воспользуемся сборкой *Debian* для различных архитектур [14], которая использовалась ранее для тестирования программного средства. Также сделаем некоторое упрощение, а именно, опустим идентификацию архитектуры *mipsel*, поскольку она достаточно близка к архитектуре *mips* и будет вносить неоправданные (и уже показанные во 2-ой части цикла [7]) неточности, как в саму матрицу, так и в вычисляемые показатели. Результат применения режима *MetricMode* представлен в таблице 1; число файлов каждой из архитектур составило 1554.

ТАБЛИЦА 1. Матрица ошибок при идентификации машинного кода файлов сборки Debian

TABLE 1. Error Matrix for Machine Code Identification of Debian Build Files

Идентифицируемые архитектуры процессора	Реальные архитектуры процессора								
	amd64	*arm64*	*armel*	*armhf*	*i386*	*mips*	*mips64el*	*ppc64el*	*s390x*
amd64	1147	5	0	5	43	5	0	6	0
arm64	0	1449	0	7	0	0	0	0	0
armel	0	0	1552	294	0	0	0	0	0
armhf	1	0	2	1066	1	0	0	0	0
i386	223	0	0	4	1421	0	0	0	0
mips	1	9	0	10	3	1319	26	0	0
mips64el	0	0	0	7	0	224	1528	0	0
ppc64el	182	91	0	160	86	6	0	1548	0
s390x	0	0	0	1	0	0	0	0	1554

Примечание. Зеленым фоном в таблице помечено наибольшее количество файлов в каждой строке, желтым – количество файлов более 10, серым – количество менее 10 (кроме 0).

Как хорошо видно (см. таблицу 1), все архитектуры были определены корректно – диагональ матрицы, указывающая на количество совпадения реальных архитектур с идентифицированными (зеленый фон), содержит максимальные значения. Однако, все же некоторое количество файлов было идентифицировано ошибочно (серый и желтый фон – в зависимости от превышения порога в 10).

Точность, полнота и F-мера

Используя матрицу ошибок, вычислим точность и полноту для идентификации каждой из архитектур. Точность (*Precession*) может быть получена, как отношение значения в диагональном элементе к сумме значений в соответствующей строке:

$$Precission_x = \frac{M_{x,x}}{\sum_{k=1..N} M_{k,x}},$$

где x – порядковый номер архитектуры; $M_{i,j}$ – значения матрицы ошибок на пересечении i-го столбца и j-ой строки; N – размер матрицы (т. е. число архитектур).

Аналогичным образом, полнота (*Recall*) может быть получена, как отношение диагонального элемента к сумме значений в столбце следующим образом:

$$Recall_x = \frac{M_{x,x}}{\sum_{k=1..N} M_{x,k}}.$$

Используя данные показатели качества, *F-мера* ($F_{measure}$) получается следующим образом:

$$F_{measure} = 2 \times \frac{Precession \times Recall}{Precession + Recall}. \quad (1)$$

Тривиальные вычисления позволяют получить следующие значения для точности, полноты и *F-меры* архитектур процессора (таблица 2).

Анализ полученных показателей (см. таблицу 2) демонстрирует, что большинство из них входят в типовой диапазон требований – больше 0,9, что является свидетельством достаточно хорошего качества работы способа и программного средства. *F-мера* считается универсальным показателем работы классификаций (включая идентификацию), и поэтому далее для определения границ применимости способа будем использовать именно ее. Предельное же значение точности, полноты и *F-меры* (т. е. 1,0) при идентификации архитектуры достигается только для архитектуры *s390x*, что говорит о ее отличительной уникальности среди остальных.

2. Границы применимости

Оценим границы применимости способа для различных условий, определяя статическое или динамическое значение *F-меры*.

Отсутствие заголовка

Для проверки работоспособности способа в режиме, не учитывающем заголовки файлов (как ELF [15], так и [16], что, соответственно, ведет к построению сигнатур не только по секциям выполняемого кода, но и по секциям данных и служебной информации), необходимо запустить средство с первым аргументом *MetricModeX*. Полученные результаты представлены в таблице 3.

Исследование условия без чтения заголовка файлов показало, что в общем случае способ имеет низкую работоспособность. Так, бóльшая часть файлов была идентифицирована, как относящаяся к архитектуре процессора *ppc64el* (серый фон), что, естественно, не верно. Это также подтверждается низким значением показателей качества – *F-мера* не превышает 0,5, при этом большинство архитектур имеет ее значение менее 0,3. Исключением, помимо *ppc64el*, является архитектура *mips* (см. соответствующий столбец таблицы 3), поскольку 2/3 реальных файлов этой архитектуры все же будет отнесено к верной архитектуре. Впрочем, в каждой строке таблицы диагональные элементы также имеют максимальное значение (зеленый фон), что говорит о том, что отдельно точность идентификации все также остается высокой.

Удаление из тестирования архитектуры *ppc64el* не повышает результативность идентификации остальных, поскольку тогда основная масса файлов начинает относиться к *mips64el*, затем к *mips* и т. д. Данный эффект возникает по причине преобладания в байтах файла (как области кода, так и данных) числа 0, что вносит сильный «шум» в процесс идентификации.

ТАБЛИЦА 2. Показатели качества идентификации машинного кода файлов сборки Debian

TABLE 2. Debian Build Files Machine Code Identification Quality Indicators

Показатели	Архитектуры процессора								
	amd64	*arm64*	*armel*	*armhf*	*i386*	*mips*	*mips64el*	*ppc64el*	*s390x*
Точность	0,95	1,00	0,84	1,00	0,86	0,96	0,87	0,75	1,00
Полнота	0,74	0,93	1,00	0,69	0,91	0,85	0,98	1,00	1,00
F-мера	0,83	0,96	0,91	0,81	0,89	0,90	0,92	0,85	1,00

Примечание. Зеленым фоном в таблице помечены ячейки со значением больше 0,9, желтым – больше 0,8, синим – больше 0,7, красным – все остальные

ТАБЛИЦА 3. Матрица ошибок и показателей качества при идентификации машинного кода файлов сборки Debian без учета заголовков файлов

TABLE 3. Error and Quality Indicators Matrix for Machine Code Identification of Debian Build Files without Regard to File Headers

Идентифицируемые архитектуры процессора	Реальные архитектуры процессора									Показатели качества		
	amd64	*arm64*	*armel*	*armhf*	*i386*	*mips*	*mips64el*	*ppc64el*	*s390x*	Точность	Полнота	*F-мера*
amd64	60	0	0	0	15	0	0	0	0	0,80	0,04	0,07
arm64	0	43	1	0	0	0	0	0	0	0,98	0,03	0,05
armel	0	0	186	0	0	0	0	0	0	1,00	0,12	0,21
armhf	1	1	2	5	1	2	1	0	1	0,36	0,00	0,01
i386	13	4	8	6	116	4	8	1	10	0,68	0,07	0,13
mips	5	8	326	76	9	962	899	2	115	0,40	0,62	0,49
mips64el	1	0	0	8	1	1	310	0	0	0,97	0,20	0,33
ppc64el	1474	1498	1031	1459	1412	585	336	1551	1287	0,15	1,00	0,25
s390x	0	0	0	0	0	0	0	0	14	1,00	0,09	0,17

Примечание. Зеленым фоном в таблице помечено наибольшее количество файлов в каждой строке, иначе серым – наибольшее количество файлов в каждом столбце.

Размер кода

Крайне интересным, как с научной, так и с практической точки зрения, является исследование зависимости *F-меры* от размера идентифицируемого МК, поскольку это покажет, насколько способ работоспособен для отдельных (небольших) участков кода. Для проведения такого эксперимента средство было временно модифицировано так, чтобы циклически брать лишь определенное количество байт из секции кода, проводить их идентификацию и получать матрицу ошибок (из которой потом вычислялись показатели качества). Результаты в виде графика зависимости *F-меры* каждой из архитектур от размера МК показаны на рисунке 1 (шаг приращения длины кода равен 50 байт).

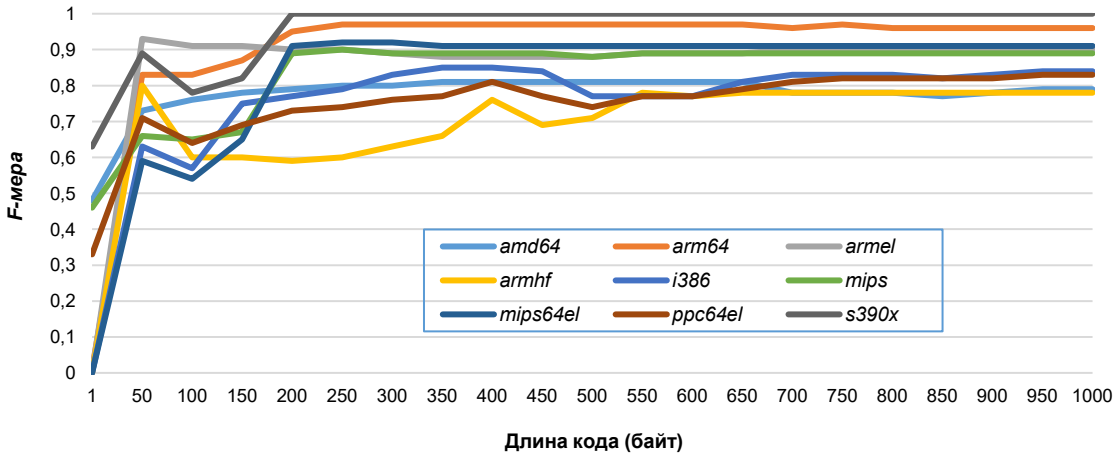


Рис. 1. Зависимость *F-меры* идентификации архитектур машинного кода от его размера

Fig. 1. The F-Measure Identification Dependence of Machine Code Architectures on its Size

На рисунке 1 в качестве оси абсцисс выбрано число байт в секции кода, по которым производилась идентификация архитектуры МК, а в качестве оси ординат – *F-мера*. При этом, поскольку вычисление *F-меры* для нулевой длины не имеет смысла, то в качестве точки отсчета по горизонтали выбран 1 байт. Также, для ряда архитектур (*armel*, *armhf*, *i386* и *mips64el*) значение точности и полноты оказалось равным нулю (при единичной длине), и, следовательно, *F-мера* по формуле (1) не может быть вычислена; в этом случае, мера принималась равной нулю.

Анализ рисунка 1 позволяет сделать вывод, что после длины секции кода примерно в 1000 байт *F-меры* достигают своих, близких к типовым, значений, указанных в таблице 2. Следовательно, данное значение можно считать нижней границей, которая позволяет получить удовлетворительные результаты работы предложенного способа идентификации.

Также, достаточно интересная картина на графике наблюдается в области длины кода 50 и 100 байт. В первом случае *F-мера* для большинства архитектур имеет ярко выраженный локальный максимум, а во втором – локальный минимум. Затем картина нормализуется, и *F-мера* начинает достаточно равномерно приближаться к типовым значениям. Такое поведение можно объяснить

тем, что на малых длинах МК происходит «конкуренция» среди различных архитектур за принадлежность к ним секций кода. При этом, данных оказывается недостаточно даже для теоретического построения полноценной сигнатуры (которая состоит из распределения 256 байт). Таким образом, вследствие ошибок идентификации на области абсциссы до 250 байт график имеет существенные колебания большинства *F-мер*.

Разрушение кода

В случае, если хранилище информации, содержащее выполняемые файлы, было подвержено частичному разрушению [17] (например, физическим воздействием на HDD или SSD, или же программным воздействием на файловую систему), идентификация архитектуры МК будет затруднена. Во-первых, если будут уничтожены заголовки файлов, то, как показали предыдущие исследования (см. таблицу 3), это может катастрофически сказаться на результативности способа. Во-вторых, если заголовки файлов не были затронуты, то МК в секциях файлов будет содержать «шум» в виде случайных байт – эффект разрушения. Такая ситуация представляет особый интерес с точки зрения определения границ применимости способа для частично разрушенных файлов. Для этого средство идентификации было временно модифицировано добавлением алгоритма, производящего замену определенного количества случайных байт МК на произвольные значения – имитируя тем самым различные степени разрушения файла. Из всех файлов бралась часть МК размером 1000 байт ($Length_{All}$, *от англ.* «длина всего»), поскольку, как показали предыдущие эксперименты (см. рисунок 1), этого размера достаточно, чтобы выйти на типовые значения *F-меры*. В качестве же количества заменяемых байт выбирались значения от 0 до 1000 ($Length_{Noize}$, *от англ.* «длина шума») с шагом 50, что позволило динамически оценить влияние разрушения файла на идентификацию:

$$K_{Destroy} = \frac{Length_{Noice}}{Length_{All}} \in [0,1].$$

Результатом исследования стала зависимость *F-меры* каждой из идентификаций архитектур МК в процессе его последовательного разрушения (рисунок 2).

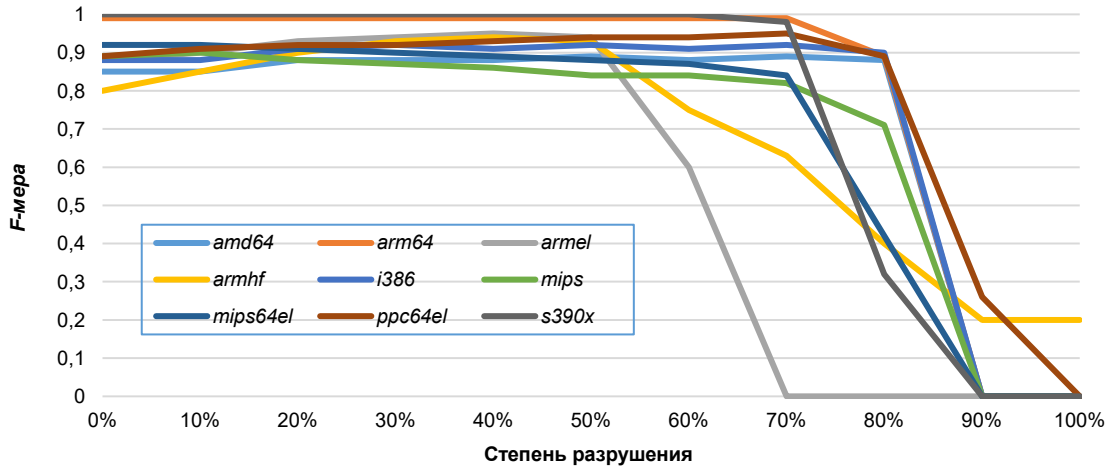


Рис. 2. Зависимость *F-меры* идентификации архитектур машинного кода от степени его разрушения

Fig. 2. The F-Measure Identification Dependence of Machine Code Architectures on the Degree of its Destruction

На рисунке 2 в качестве оси абсцисс выбрана $K_{Destroy}$ – степень разрушения (*Destroy*), выраженная в процентах, а в качестве оси ординат – *F-мера*. При этом, поскольку для ряда измерений точность и полнота оказались равными 0 (область более 70 % по оси абсциссы), то для них *F-мера* принималась равной 0.

Анализ полученных результатов (см. рисунок 2) позволяет сделать следующие выводы. Во-первых, значение *F-меры* для большинства архитектур скачкообразно падает с 50 до 90 % разрушения МК файла; при этом условно критической точкой можно считать 70 %. Во-вторых, при 100 %-ном разрушении (состояние «белого шума») любой МК идентифицируется, как *armhf*. И, в-третьих, на всем процессе разрушения *F-мера* архитектур испытывает колебание в обе стороны – т. е. отсутствует строго снижающаяся динамика. Последнее видимо объяснимо особенностями распределения байт в МК различных архитектур процессора (см. рисунок 1 первой части цикла статей [6]), а точнее, их близости к случайному распределению. Также, частотная сигнатура *armhf* визуально имеет меньшее количество ярко выраженных пиков (или большую площадь распределения), что, видимо и влияет на то, что МК с преобладающим количеством случайных байт оказывается более близок именно к этой архитектуре.

Код разных архитектур

В практике крупного телекоммуникационного оборудования возможна ситуация, когда в одном МК содержатся инструкции сразу нескольких процессорных архитектур; например, присутствует

как загрузчик, выполняемый на основном оборудовании, так и код прошивки периферийного устройства другой архитектуры. Следовательно, необходимо исследовать работу способа в такой ситуации. Для этого можно воспользоваться простым экспериментом, заключающимся не в вычислении сигнатур множества файлов с кодом двух архитектур, а в сложении частотных сигнатур двух классов (архитектур) МК – т. е. тех, которые используются для машинного обучения. Результаты идентификации файлов, которым соответствуют попарные комбинации сигнатур различных архитектур, представлены в таблице 4; последней строкой в таблице идет комбинация всех девяти архитектур.

ТАБЛИЦА 4. Вероятности отнесения комбинированных файлов к архитектурам процессоров

TABLE 4. Probabilities of Attributing Combined Files to Processor Architectures

Тестовые файлы		Архитектуры процессора								
Архитектура 1	*Архитектура 2*	*amd64*	*arm64*	*armel*	*armhf*	*i386*	*mips*	*mips64el*	*ppc64el*	*s390x*
amd64	*arm64*	0,26	0,13	0,07	0,07	0,06	0,10	0,10	0,12	0,08
amd64	*armel*	0,22	0,08	0,12	0,11	0,07	0,10	0,10	0,11	0,10
amd64	*armhf*	0,20	0,09	0,11	0,13	0,08	0,09	0,10	0,10	0,10
amd64	*i386*	0,15	0,08	0,07	0,07	0,37	0,06	0,07	0,07	0,07
amd64	*mips*	0,13	0,06	0,05	0,06	0,09	0,30	0,12	0,11	0,08
amd64	*mips64el*	0,28	0,08	0,07	0,07	0,06	0,12	0,13	0,11	0,08
amd64	*ppc64el*	0,15	0,06	0,06	0,06	0,09	0,07	0,07	0,37	0,08
amd64	*s390x*	0,13	0,08	0,06	0,06	0,10	0,09	0,09	0,11	0,28
arm64	*armel*	0,09	0,13	0,24	0,12	0,07	0,09	0,09	0,09	0,09
arm64	*armhf*	0,10	0,24	0,11	0,13	0,07	0,08	0,09	0,09	0,09
arm64	*i386*	0,12	0,13	0,08	0,08	0,21	0,09	0,09	0,11	0,10
arm64	*mips*	0,12	0,13	0,06	0,06	0,06	0,29	0,11	0,10	0,08
arm64	*mips64el*	0,12	0,13	0,06	0,06	0,06	0,08	0,30	0,10	0,08
arm64	*ppc64el*	0,13	0,27	0,06	0,06	0,07	0,10	0,10	0,13	0,08
arm64	*s390x*	0,08	0,13	0,07	0,07	0,07	0,11	0,11	0,11	0,26
armel	*armhf*	0,09	0,11	0,35	0,14	0,06	0,06	0,06	0,06	0,07
armel	*i386*	0,12	0,08	0,18	0,10	0,12	0,09	0,09	0,10	0,11
armel	*mips*	0,09	0,11	0,23	0,07	0,07	0,12	0,11	0,11	0,09
armel	*mips64el*	0,09	0,11	0,20	0,08	0,08	0,12	0,12	0,11	0,09
armel	*ppc64el*	0,10	0,11	0,12	0,11	0,06	0,08	0,07	0,26	0,09
armel	*s390x*	0,07	0,07	0,14	0,12	0,06	0,08	0,08	0,08	0,30
armhf	*i386*	0,07	0,08	0,10	0,12	0,20	0,10	0,10	0,11	0,11
armhf	*mips*	0,08	0,10	0,11	0,11	0,06	0,21	0,11	0,10	0,11
armhf	*mips64el*	0,09	0,10	0,11	0,19	0,07	0,11	0,12	0,10	0,11
armhf	*ppc64el*	0,10	0,11	0,10	0,12	0,06	0,06	0,07	0,24	0,12
armhf	*s390x*	0,06	0,07	0,11	0,28	0,07	0,09	0,09	0,09	0,13
i386	*mips*	0,12	0,07	0,06	0,07	0,12	0,24	0,12	0,11	0,09
i386	*mips64el*	0,12	0,07	0,06	0,07	0,24	0,11	0,12	0,11	0,09
i386	*ppc64el*	0,12	0,06	0,06	0,07	0,14	0,07	0,07	0,31	0,10
i386	*s390x*	0,11	0,08	0,06	0,07	0,13	0,08	0,09	0,10	0,28
mips	*mips64el*	0,12	0,08	0,07	0,06	0,06	0,30	0,13	0,06	0,12
mips	*ppc64el*	0,13	0,06	0,05	0,05	0,05	0,13	0,10	0,32	0,12
mips	*s390x*	0,08	0,08	0,06	0,06	0,07	0,13	0,12	0,10	0,29
mips64el	*ppc64el*	0,13	0,06	0,05	0,05	0,05	0,11	0,13	0,31	0,11
mips64el	*s390x*	0,08	0,08	0,06	0,06	0,07	0,12	0,13	0,10	0,28
ppc64el	*s390x*	0,08	0,09	0,06	0,06	0,07	0,11	0,10	0,13	0,28
Все архитектуры		0,11	0,11	0,10	0,10	0,10	0,11	0,11	0,15	0,10

Примечание. Зеленым фоном в таблице помечены наибольшие вероятности, соответствующие идентифицированным архитектурам процессора МК, желтым – ближайшая альтернатива, серым – первая корректно идентифицированная архитектура из комбинации, голубым – вторая корректно идентифицированная архитектура из комбинации.

Результаты идентификации МК, содержащего несколько архитектур процессора (см. таблицу 4) позволяют сделать следующие выводы. Во-первых, для всех комбинаций пар архитектур обе из них были определены верно, поскольку имеют максимальную и вторую по величине вероятности (в двух левых столбцах архитектура отмечена синим или голубым фоном). Заметим, что в некоторых строках желтый фон имеют сразу несколько ячеек, поскольку численные вероятности идентификации архитектур совпали; причина этого, скорее всего, кроется в ограниченном выводе знаков после запятой. Во-вторых, различия между максимальной вероятностью и второй по величине можно считать значимыми (минимально, около 30 %), и, следовательно, идентификация хотя бы одной из архитектур является однозначной. И, в-третьих, идентификация «смеси» из всех возможных архитектур (последняя строка таблицы) показала примерно одинаковое распределение вероятностей (0,10 и 0,11) с некоторым «перекосом» в сторону архитектуры *ppc64el*, что необходимо учитывать при практическом использовании.

3. Обсуждение результатов

Исследование показателей качества способа идентификации и границ его применимости позволили определить следующее.

Во-первых, *F-мера* для всех архитектур процессора превосходит 0,8, а для половины и – 0,9, что является достаточно высоким показателем. Еще большего их повышения можно достигнуть усложнением частотно-байтовой модели, например, введением в нее частоты «встречания» двух (или более) определенных байт подряд – *N*-грамм [18, 5], или использованием специальных семантик машинных инструкций [19].

Во-вторых, интерпретация выполняемых файлов, как последовательности байт без учета их заголовков (и, следовательно, без разбиения на секции кода, данных и служебную информацию) показывает низкую работоспособность из-за отнесения большинства таких файлов исключительно к одной из архитектур. Впрочем, тот факт, что отдельно показатель точности остается высоким, позволяет предположить, что модернизация способа позволит успешно его применять и в этом случае (аналогично работам по поиску упакованных PE-файлов [20–22]), в том числе противодействуя сложным системам сокрытия информации [23].

В-третьих, разрушение 70 % и более МК путем замены некоторого количества байт на случайные в конечном итоге приводит к неверной идентификации архитектуры процессора, как *armhf*, вместо ожидаемого равномерного распределения среди архитектур. Такое поведение отражает особенности набора инструкций МК рассматриваемых архитектур и способов их байтового кодирования [24]; поэтому, повышение результативности такой идентификации представляется мало реалистичным.

И, в-четвертых, применение средства для выполняемых файлов, МК которых содержит инструкции сразу двух процессорных архитектур, показало высокую работоспособность – обе архитектуры идентифицируются приемлемым образом; это, очевидно, расширяет область применения способа; например, для поиска дублируемого функционала [25].

Заключение

Таким образом, в третьей части цикла проведены полноценное тестирование и количественная оценка способа и средства идентификации с определением соответствующих показателей качества: точности, полноты и *F-меры*. Последние показали высокие значения, доказывающие применимость способа на практике; в том числе, в качестве предварительного этапа для авторского *метода алгоритмизации* [26, 27] (преобразующего МК в вид, подходящий эксперту для поиска уязвимостей), а также при начальном этапе интеллектуального статического анализа [28]. Были определены границы применимости способа для наиболее типовых условий работы: отсутствие заголовков файлов, влияние размера и степени разрушения МК, наличие инструкций процессора нескольких архитектур. Несмотря на ограниченность применимости в некоторых случаях, предполагается, что дальнейшая и более «тонкая» доработка способа позволит частично устранить эти недостатки.

Основное развитие способа, с точки зрения авторов, должно быть направлено именно на обеспечение его работоспособности в условиях отсутствия заголовков файлов, которые позволяют однозначно выделить секции с инструкциями МК из другой информации. Данное усовершенствование способа может быть осуществлено внедрением предварительного этапа выделения областей с МК, в том числе с применением интеллектуальных методов выявления аномалий. В этом случае будет возможным поиск и точная идентификация процессорной архитектуры участков с выполняемым кодом в любых бинарных областях независимо от окружающего контекста [29].

Окончание цикла.

Список используемых источников

1. Буйневич М.В., Васильева И.Н., Воробьев Т.М., Гниденко И.Г., Егорова И.В. и др. Защита информации в компьютерных системах: монография. СПб.: Санкт-Петербургский государственный экономический университет, 2017. 163 с.

2. Kim J., Youn J.M. Malware behavior analysis using binary code tracking // Proceedings of the 4th International Conference on Computer Applications and Information Processing Technology (CAIPT, Kuta Bali, Indonesia, 8–10 August 2017). IEEE, 2017. DOI:10.1109/CAIPT.2017.8320724

3. Elhadi A.A.E., Maarof M.A., Barry B.I.A. Improving the Detection of Malware Behaviour Using Simplified Data Dependent API Call Graph // International Journal of Security and Its Applications. 2013. Vol. 7. Iss. 5. PP. 29–42. DOI:10.14257/ijsia.2013.7.5.03

4. Anwar Z., Sharf M., Khan E., Mustafa M. VG-MIPS: A dynamic binary instrumentation framework for multi-core MIPS processors // Proceedings of the International Conference on Multi Topic (INMIC, Lahore, Pakistan, 19–20 December 2013). 2013. IEEE, 2013. PP. 166–171. DOI:10.1109/INMIC.2013.6731344

5. Erozan A.S.A. File fragment type detection by neural network // Proceedings of the 26th Signal Processing and Communications Applications Conference (SIU, Izmir, Turkey, 2–5 May 2018). IEEE, 2018. DOI:10.1109/SIU.2018.8404380

6. Буйневич М.В., Израилов К.Е. Идентификация архитектуры процессора выполняемого кода на базе машинного обучения. Часть 1. Частотно-байтовая модель // Труды учебных заведений связи. 2020. Т. 6. № 1. С. 77–85. DOI:10.31854/1813-324X-2020-6-1-77-85

7. Буйневич М.В., Израилов К.Е. Идентификация архитектуры процессора выполняемого кода на базе машинного обучения. Часть 2. Способ идентификации // Труды учебных заведений связи. 2020. Т. 6. № 2. С. 104–112. DOI:10.31854/1813-324X-2020-6-2-104-112

8. Шунина Ю.С., Алексеева В.А., Клячкин В.Н. Критерии качества работы классификаторов // Вестник Ульяновского государственного технического университета. 2015. № 2(70). С. 67–70.

9. Трофименков А.К., Трофименков С.А., Пимонов Р.В. Алгоритмизация обработки файлов для их идентификации при нарушении целостности данных // Системы управления и информационные технологии. 2020. № 2(80). С. 82–85.

10. Антонов А.Е., Федулов А.С. Идентификация типа файла на основе структурного анализа // Прикладная информатика. 2013. № 2(44). С. 068–077.

11. Касперски К. Как спасти данные, если отказал жесткий диск // Системный администратор. 2005. № 9(34). С. 80–87.

12. Кажемский М.А., Шелухин О.И. Многоклассовая классификация сетевых атак на информационные ресурсы методами машинного обучения // Труды учебных заведений связи. 2019. Т. 5. № 1. С. 107–115. DOI:10.31854/1813-324X-2019-5-1-107-115

13. Попков М.И. Автоматическая система классификации текстов для базы знаний предприятия // International Journal of Open Information Technologies. 2014. Т. 2. № 7. С. 11–18.

14. Файлы образов Debian версии 10.3.0 // Debian. URL: https://www.debian.org/distrib/netinst.ru.html (дата обращения 20.03.2020)

15. Штеренберг С.И., Красов А.В. Варианты применения языка ассемблера для заражения вирусом исполнимого файла формата ELF // Информационные технологии и телекоммуникации. 2013. Т. 1. № 3. С. 61–71.

16. Юрин И.Ю. Способы установления первоначального имени PE-файла // Теория и практика судебной экспертизы. 2008. № 3(11). С. 200–205.

17. Жилин В.В., Сафарьян О.А. Искусственный интеллект в системах хранения данных // Вестник Донского государственного технического университета. 2020. Т. 20. № 2. С. 196–200. DOI:10.23947/1992-5980-2020-20-2-196-200

18. Al-Kasassbeh M., Mohammed S., Alauthman M., Almomani A. Feature Selection Using a Machine Learning to Classify a Malware // Gupta B., Perez G., Agrawal D., Gupta D. (eds) Handbook of Computer Networks and Cyber Security. Springer, Cham, 2020. PP. 889–904. DOI:10.1007/978-3-030-22277-2_36

19. Падарян В.А., Соловьев М.А., Кононов А.И. Моделирование операционной семантики машинных инструкций // Труды Института системного программирования РАН. 2010. Т. 19. С. 165–186.

20. Wang T.-Y., Wu C.-H. Detection of packed executables using support vector machines // Proceedings of the International Conference on Machine Learning and Cybernetics (Guilin, China, 10–13 July 2011). IEEE, 2011. PP. 717–722. DOI:10.1109/ICMLC.2011.6016774

21. Hubballi N., Dogra H. Detecting Packed Executable File: Supervised or Anomaly Detection Method? // Proceedings of the 11th International Conference on Availability, Reliability and Security (ARES, Salzburg, Austria, 31 August–2 September 2016). IEEE, 2016. PP. 638–643. DOI:10.1109/ARES.2016.18

22. Choi Y.-S., Kim I.-K., Oh J.-T., Ryou J.-C. PE File Header Analysis-Based Packed PE File Detection Technique (PHAD) // Proceedings of the International Symposium on Computer Science and its Applications (Hobart, Australia, 13–15 October 2008). IEEE, 2008. PP. 28–31. DOI:10.1109/CSA.2008.28

23. AL-Nabhani Y., Zaidan A.A., Zaidan B.B., Jalab H.A., Alanazi H.O. A new system for hidden data within header space for EXE-File using object oriented technique // Proceedings of the 3rd International Conference on Computer Science and Information Technology (Chengdu, China, 9–11 July 2010). IEEE, 2010. PP. 9–13. DOI:10.1109/ICCSIT.2010.5564461

24. Соловьев М.А., Бакулин М.Г., Макаров С.С., Манушин Д.В., Падарян В.А. Декодирование машинных команд в задаче абстрактной интерпретации бинарного кода // Труды Института системного программирования РАН. 2019. Т. 31. № 6. С. 65–88. DOI:10.15514/ISPRAS-2019-31(6)-4

25. Wang M., Tang Y., Lu Z. Massive Similar Function Searching for Cross-Architecture Binaries // Proceedings of the 18th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES, Wuhan, China, 8–10 November 2019). IEEE, 2019. PP. 195–198. DOI:10.1109/DCABES48411.2019.00055

26. Буйневич М.В., Израилов К.Е. Метод алгоритмизации машинного кода телекоммуникационных устройств // Телекоммуникации. 2012. № 12. С. 2–6.

27. Буйневич М.В., Израилов К.Е. Автоматизированное средство алгоритмизации машинного кода телекоммуникационных устройств // Телекоммуникации. 2013. № 6. С. 2–9.

28. Буйневич М.В., Израилов К.Е. Обобщенная модель статического анализа программного кода на базе машинного обучения применительно к задаче поиска уязвимостей // Информатизация и Связь. 2020. № 2. С. 143–152. DOI:10.34219/2078-8320-2020-11-2-143-152

29. Поддубный В.А., Коркин И.Ю. Средство обнаружения скрытого исполнимого кода в памяти ОС Windows // Вопросы кибербезопасности. 2019. № 5(33). С. 75–82. DOI:10.21681/2311-3456-2019-5-75-82

* * *

Identification of Processor's Architecture of Executable Code Based on Machine Learning. Part 3. Assessment Quality and Applicability Border

M. Buinevich[1, 2], **K. Izrailov**[1, 3]

[1]The Bonch-Bruevich Saint-Petersburg State University of Telecommunications,
St. Petersburg, 193232, Russian Federation
[2]Saint-Petersburg University of State Fire Service of Emercom of Russia,
St. Petersburg, 195105, Russian Federation
[3]St. Petersburg Federal Research Center of the Russian Academy of Sciences,
St. Petersburg, 199178, Russian Federation

Article info
DOI:10.31854/1813-324X-2020-6-3-48-57
Received 10th July 2020
Accepted 8th September 2020

For citation: Buinevich M., Izrailov K. Identification of Processor's Architecture of Executable Code Based on Machine Learning. Part 3. Assessment Quality and Applicability Border. *Proc. of Telecom. Universities.* 2020;6(3): 48–57. (in Russ.) DOI:10.31854/1813-324X-2020-6-3-48-57

Abstract: *The article presents the author's method testing results for identifying the processor architecture of the executable code based on machine learning. In the third final part of the cycle, its qualitative indicators are determined: accuracy, completeness and F-measure for the executable files of the Debian build. There are investigated the applicability limits of the architecture identification method for four conditions: the file header absence, different sizes of machine code, partial code destruction, and the presence of instructions from several architectures. We can observe the identified disadvantages of the proposed method and ways to eliminate them, as well as the further direction of its development.*

Keywords: *information security, machine code, processor architecture, machine learning, code signature, identification method, quality indicators, error matrix, accuracy, completeness, F-measure, code destruction*

References

1. Buynevich M.V., Vasilieva I.N., Vorobyov T.M., Gnidenko I.G., Egorova I.V. et al. *Information Security in Computer Systems*. St. Petersburg: Saint Petersburg Electrotechnical University "LETI" Publ.; 2017. 163 p. (in Russ.)

2. Kim J., Youn J.M. Malware behavior analysis using binary code tracking. *Proceedings of the 4th International Conference on Computer Applications and Information Processing Technology, CAIPT, 8–10 August 2017, Kuta Bali, Indonesia*. IEEE; 2017. DOI:10.1109/CAIPT.2017.8320724

3. Elhadi A.A.E., Maarof M.A., Barry B.I.A. Improving the Detection of Malware Behaviour Using Simplified Data Dependent API Call Graph. *International Journal of Security and Its Applications*. 2013;7(5):29–42. DOI:10.14257/ijsia.2013.7.5.03

4. Anwar Z., Sharf M., Khan E., Mustafa M. VG-MIPS: A dynamic binary instrumentation framework for multi-core MIPS processors. *Proceedings of the International Conference on Multi Topic, INMIC, 19–20 December 2013, Lahore, Pakistan*. 2013. IEEE; 2013. p.166–171. DOI:10.1109/INMIC.2013.6731344

5. Erozan A.S.A. File fragment type detection by neural network. *Proceedings of the 26th Signal Processing and Communications Applications Conference, SIU, 2–5 May 2018, Izmir, Turkey*. IEEE; 2018. DOI:10.1109/SIU.2018.8404380

6. Buinevich M., Izrailov K. Identification of Processor's Architecture of Executable Code Based on Machine Learning. Part 1. Frequency Byte Model. *Proc. of Telecom. Universities*. 2020;6(1):77–85. (in Russ.) DOI:10.31854/1813-324X-2020-6-1-77-85

7. Buinevich M., Izrailov K. Identification of Processor's Architecture of Executable Code Based on Machine Learning. Part 2. Identification method. *Proc. of Telecom. Universities.* 2020;6(2):104–112. (in Russ.) DOI:10.31854/1813-324X-2020-6-2-104-112

8. Shunina Ju. S., Alekseeva V.A., Klyachkin V.N. Criteria of Quality of Qualifiers Work. *Vestnic of UlSTU*. 2015;2(70):67–70. (in Russ.)

9. Trofimenkov A.K., Trofimenkov S.A., Pimonov R.V. Algorithmization of File Processing for their Identification in Case of Violation of Data Integrity. *Sistemy upravleniya i informatsionnyye tekhnologii*. 2020;2(80):82–85. (in Russ.)
10. Antonov A., Fedulov A. File type identification based on structural analyses. *Journal of Applied Informatics*. 2013;2(44):068–077. (in Russ.)
11. Kaspersky K. How to Save Data if the Hard Drive Fails. *Sistemnyy administrator*. 2005;9(34):80–87.
12. Kazhemskiy M., Sheluhin O. Multiclass Classification of Attacks to Information Resources with Machine Learning Techniques. *Proc. of Telecom. Universities*. 2019;5(1):107–115. (in Russ.) DOI:10.31854/1813-324X-2019-5-1-107-115
13. Popkov M.I. Text Analytics for Enterprise Knowledge Base. *International Journal of Open Information Technologies*. 2014;2(7):11–18. (in Russ.)
14. *Debian.* Debian Image Files Version 10.3.0. Available from: https://www.debian.org/distrib/netinst.ru.html (in Russ.) [Accessed 20th March 2020]
15. Shterenberg S.I., Krasov A.V. Methods of Using Assembly Language for Infection the Virus Executable File Format ELF. *TelecomIT*. 2013;1(3):61–71. (in Russ.)
16. Yurin I.Yu. Ways to Set the Initial PE File Name. *Theory and Practice of Forensic Science*. 2008;3(11):200–205. (in Russ.)
17. Zhilin V.V., Safar'yan O.A. Rtificial Intelligence in Data Storage Systems. *Vestnik of Don State Technical University*. 2020; 20(2):196–200. DOI:10.23947/1992-5980-2020-20-2-196-200
18. Al-Kasassbeh M., Mohammed S., Alauthman M., Almomani A. Feature Selection Using a Machine Learning to Classify a Malware. *In: Gupta B., Perez G., Agrawal D., Gupta D. (eds) Handbook of Computer Networks and Cyber Security*. Springer, Cham; 2020. p.889–904. DOI:10.1007/978-3-030-22277-2_36
19. Padaryan V.A., Soloviev M.A., Kononov A.I. Modeling the operational semantics of machine instructions. *Proceedings of the Institute for System Programming of the RAS*. 2010;19:165–186. (in Russ.)
20. Wang T.-Y., Wu C.-H. Detection of packed executables using support vector machines. *Proceedings of the International Conference on Machine Learning and Cybernetics 10–13 July 2011, Guilin, China*. IEEE; 2011. p.717–722. DOI:10.1109/ICMLC.2011.6016774
21. Hubballi N., Dogra H. Detecting Packed Executable File: Supervised or Anomaly Detection Method? *Proceedings of the 11th International Conference on Availability, Reliability and Security, ARES, 31 August–2 September 2016, Salzburg, Austria*. IEEE; 2016. p.638–643. DOI:10.1109/ARES.2016.18
22. Choi Y.-S., Kim I.-K., Oh J.-T., Ryou J.-C. PE File Header Analysis-Based Packed PE File Detection Technique (PHAD). *Proceedings of the International Symposium on Computer Science and its Applications, 13–15 October 2008, Hobart, Australia*. IEEE; 2008. p.28–31. DOI:10.1109/CSA.2008.28
23. AL-Nabhani Y., Zaidan A.A., Zaidan B.B., Jalab H.A., Alanazi H.O. A new system for hidden data within header space for EXE-File using object oriented technique. *Proceedings of the 3rd International Conference on Computer Science and Information Technology, 9–11 July 2010, Chengdu, China*. IEEE; 2010. p.9–13. DOI:10.1109/ICCSIT.2010.5564461
24. Solovev M.A., Bakulin M.G., Makarov S.S., Manushin D.V., Padaryan V.A. Decoding of Machine Instructions for Abstract Interpretation of Binary Code. *Proceedings of the Institute for System Programming of the RAS*. 2019;31(6):65–88. DOI: 10.15514/ISPRAS-2019-31(6)-4 (in Russ.)
25. Wang M., Tang Y., Lu Z. Massive Similar Function Searching for Cross-Architecture Binaries. *Proceedings of the 18th International Symposium on Distributed Computing and Applications for Business Engineering and Science, DCABES, 8–10 November 2019, Wuhan, China*. IEEE; 2019. p.195–198. DOI:10.1109/DCABES48411.2019.00055
26. Buinevich M.V., Izrailov K.E. Algorithmization Method for Machine Code of Telecommunication Devices. *Telekommunikatsii (Telecommunications)*. 2012;12:2–6. (in Russ.)
27. Buinevich M.V., Izrailov K.E. Automated Tool for Algorithmic Machine Code of Telecommunication Devices. *Telekommunikatsii (Telecommunications)*. 2013;6:2–9. (in Russ.)
28. Buinevich M., Izrailov K. A Generalized Model of Static Analysis of Program Code Based on Machine Learning for the Vulnerability Search Problem. *Informatizatsiya i Svyaz'*. 2020;2:143–152 (in Russ.) DOI:10.34219/2078-8320-2020-11-2-143-152
29. Poddubnyy V., Korkin I. Advanced Rootkit Detection Using Memory Forensics. Voprosy Kiberbezopasnosti. 2019;5(33): 75–82. DOI:10.21681/2311-3456-2019-5-75-82 (in Russ.)

Сведения об авторах:

БУЙНЕВИЧ Михаил Викторович — доктор технических наук, профессор, профессор кафедры безопасности информационных систем Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича, профессор кафедры прикладной математики и информационных технологий Санкт-Петербургского университета государственной противопожарной службы МЧС России, bmv1958@yandex.ru
https://orcid.org/0000-0001-8146-0022

ИЗРАИЛОВ Константин Евгеньевич — кандидат технических наук, доцент кафедры защищенных систем связи Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича, старший научный сотрудник Санкт-Петербургского федерального исследовательского центра Российской академии наук, konstantin.izrailov@mail.ru
https://orcid.org/0000-0002-9412-5693

УДК 004.891.2 DOI:10.31854/1813-324X-2020-6-3-58-65

Метод валидации графовых моделей на основе алгоритма эффективных управлений

В.С. Васильев[1], **А.Н. Целых**[1*], **Л.А. Целых**[2]

[1]Институт компьютерных технологий и информационной безопасности Южного федерального университета, Таганрог, 347900, Российская Федерация
[2]Таганрогский институт имени А.П. Чехова (филиал) Ростовского государственного экономического университета, Таганрог, 347936, Российская Федерация
*Адрес для переписки: ant@sfedu.ru

Информация о статье
Поступила в редакцию 01.08.2020
Принята к публикации 19.08.2020

Ссылка для цитирования: Васильев В.С., Целых А.Н., Целых Л.А. Метод валидации графовых моделей на основе алгоритма эффективных управлений // Труды учебных заведений связи. 2020. Т. 6. № 3. С. 58–65. DOI:10.31854/1813-324X-2020-6-3-58-65

Аннотация: *В статье предлагается метод валидации математических моделей, представленных ориентированными взвешенными знаковыми графами, с использованием алгоритма эффективных управлений. Метод рассматривает валидируемую модель с точки зрения спектральных свойств матрицы смежности графа, представленного нечеткой когнитивной картой (НКК). Использование алгоритма эффективных управлений позволяет определить направленность собственного вектора матрицы смежности. Это свойство определяет критерии проверки НКК.*

Ключевые слова: *валидация моделей, направленный взвешенный граф, нечеткая когнитивная карта, эффективное управление.*

Введение

Сложные системы состоят из большого количества гетерогенных взаимодействующих сущностей, которые называются концептами или факторами системы. В качестве концептов могут выступать факторы внешней и внутренней среды, влияющие на работу системы. Элементы системы имеют различную природу и разные характеристики. Это могут быть элементы инфраструктуры, внешние относительно исследуемой системы, или внутренние подсистемы, например, объекты или системы управления. Сами эти системы и их отдельные компоненты имеют нелинейное поведение и относятся к классу сложных динамических систем. В этом случае традиционные методы моделирования имеют ограниченное применение.

Моделирование сложных систем требует специальных методов, которые основаны на имеющихся знаниях о функционировании системы. При этом следует учитывать такие специфические свойства предметной области, как качественный характер факторов системы, причинно-следственный характер связей, целостность системы, изменчивость системы во времени. Методология нечеткой когнитивной карты (НКК), предложенная Коско [1], получила большое распространение для моделирования таких систем. С точки зрения теории графов, НКК является ориентированным взвешенным знаковым графом и может быть представлена в виде матрицы смежности.

При использовании этой методологии возникают три вопроса: технология разработки НКК, моделирование динамических изменений, использование результатов моделирования.

Основные исследования в области конструирования НКК главным образом сосредоточены на определении весов на ребрах с применением для этого различных алгоритмов, в т. ч. нейронных сетей, эволюционных алгоритмов и др. [2]. Основным методом моделирования динамики НКК является метод векторно-матричного перемножения в автономной модели. Этот метод позволяет вычислять состояние концептов после действия входного импульса, задаваемого вектором состояния. Такой подход требует соблюдения условий устойчивости, т. е. нахождения всех собственных значений оператора (матрицы смежности графа) внутри единичного круга. Это строго необходимое условие не всегда учитывается при построении модели, что часто приводит к неустойчивости общего решения.

Цель исследования – представить метод валидации графовых моделей на основе алгоритма эффективных управлений, показать его преимущество и возможность использования для валидации и вери-

фикации моделей, представленных НКК. Предлагаемый метод валидации интегрирует алгоритм эффективных управлений в рабочий процесс валидации.

Обзор публикаций

Как отмечается во многих публикациях, исследующих НКК, проблема преодоления субъективизма при их конструировании является одной из наиболее сложных задач. Основным подходом к решению этой задачи является разработка методов обучения НКК. Последнее исследование [2], посвященное обучению НКК, ставит вопрос валидации, как одну из главных проблем НКК, характеризуя ее как «жизненно важный вопрос». Алгоритмы обучения используют следующие подходы: популяционные алгоритмы, в т. ч. искусственные нейронные сети, генетические алгоритмы и др.; алгоритмы Хебба; гибридные алгоритмы и другие. Однако надо отметить, что все подходы к валидации посредством обучения НКК ограничены необходимостью нахождения «матрицы весов, которая лучше всего подходит для конкретной задачи». Это называется «тонкой настройкой НКК».

Критерий окончания обучения НКК рассматривается как критерий окончания работы алгоритма. В качестве такового используются естественное уменьшение весов (алгоритм с сетями Петри [3]), приемлемый диапазон для целевой задачи (алгоритм обучения Хебба [4]), конечная сходимость модели (алгоритм обучения «большой взрыв – большое сжатие» [5]), оптимизация целевой функции НКК (алгоритм самоорганизующейся миграции [6]), количество итераций (модифицированная оптимизация бесполого размножения [2]), ограничения, накладываемые на узлы (нелинейное обучение Хебба [7]). В качестве целевой функции для оптимизационной задачи, в основном, используется достижение желаемых значений в изменении показателей вершин.

При этом предполагается, что если правильно определены вершины и веса на ребрах между вершинами, то верна и разработанная модель. Такой подход не рассматривает систему в целом. Поэтому требуется разработка специальной модели, в которой используется критерий валидации самой модели системы в целом, а не алгоритмов, используемых для верификации весовой матрицы.

Метод валидации графовых моделей, основанный на теории эффективных управлений

Эволюцию методов валидации математических моделей можно представить как движение от процедур с человеческим участием до автономных процедур без участия человека. Соответственно, существующие системы поддержки принятия решения (СППР) либо копируют человеческое поведение при решении задач, либо заменяют знания человека-эксперта неструктурированными артефактными данными. В работе [8] отмечаются, что такой подход не учитывает изменение восприятия полученных знаний в результате процедуры валидации, что влечет изменение собственного уровня знаний и воздействует на онтологическую конструкцию знания. Изменение взаимодействия человека-эксперта и информационной системы предопределяют новую ступень развития методов валидации. Коэволюционный подход, анонсированный этим исследованием [8], может быть поддержан сравнением некоторых характеристик методов валидации, основанных на извлечении знаний, которые представлены в таблице 1.

Приведенный анализ показывает обоснованность и возможность увеличения конверсии классических методов извлечения знаний с новыми возможностями инструментария обнаружения знаний. На наш взгляд, реализацией такой идеи может служить интеграция ментальных представлений о работе системы в целом с математическим подходом к обработке таких моделей для дальнейшего использования в СППР. Ментальные представления о системе могут быть представлены НКК, соответствующая математическая обработка которых даст обоснованные выходные данные для использования созданных моделей в СППР.

Исходя из изложенного в предыдущих разделах, структура метода должна обладать двумя важными характеристиками: системным подходом и математическим обеспечением. Как утверждалось в [2], качество базы знаний зависит от его пяти детерминант. В предлагаемой модели каждая из них получила определенное развитие, представленное в таблице 2. Предлагаемые инструменты, реализованные в рассматриваемом методе, позволяют снизить субъективизм экспертов предметной области и учитывать ее специфику.

Наш подход реализуется с помощью набора инструментов, который включает в себя: алгоритм эффективных управлений, метод обучения НКК и метод реконструкции НКК, что позволяет сформировать системное представление предметной области (исследуемой системы). Реализация метода валидации графовых моделей представлена на рисунке 1. Первичное формирование НКК не является предметом нашего исследования, поэтому процесс ее создания и связанные с этим проблемы в исследовании не отражаются.

Математическая обработка графовой модели исследуемой системы реализуется с использованием алгоритма эффективных управлений [9, 14]. Проверка модели предусматривает два этапа ее обучения: верификацию и валидацию. Первый этап, верификация модели, основан на анализе результатов ее математической обработки. Эксперт получает результат (факторы воздействия и

факторы влияния, упорядоченные по убыванию) и сверяет его со своими ментальными представлениями о работе системы в целом, используя методологию оценки продуктивности влияния, представленную в работе [15]. В случае несоответствия работы модели его представлениям об этом, эксперт запускает процедуру верификации модели с помощью процедуры «обучение НКК», т. е. начинает ее обучение. Каждый раз, сверяя полученный результат математической обработки, эксперт корректирует либо вес, либо факторы, либо причинно-следственные связи, и добивается нужной модели. Разработанная модель поступает в библиотеку темпоральных моделей, привязанных к конкретной решаемой задаче и временному периоду. Второй этап связан с изменениями работы исследуемой системы во времени, как внешними, так и внутренними. Если такие изменения наступили, пользователь, запросив модель из библиотеки, приступает к ее валидации через процедуру «реконструкция НКК». Таким образом, первый этап обучения НКК (верификация) связан с некоторыми недоработками при разработке модели, второй этап обучения (валидация) – с изменением условий существования модели. Все разработанные модели формируют библиотеку моделей и могут быть использованы для дальнейшего анализа и экспертизы.

ТАБЛИЦА 1. Сравнительная характеристика методов валидации, основанных на знаниях

TABLE 1. Comparative Characteristics of Knowledge-Based Validation Methods

Характеристики процесса извлечения знаний	Приобретение знаний (ПЗ)	Извлечение знаний (ИЗ)	ПЗ → ИЗ
Метод получения знаний	анкета, интервью	сбор данных	поддержка НКК + математическое моделирование
Тип системныжнх факторов	количественные, качественные	количественные	количественные, качественные
Факторы, характеристики системы	наблюдаемые	наблюдаемые	наблюдаемые, ненаблюдаемые
Тип знаний	явные, неявные	скрытые	явные, неявные, скрытые
Носитель знаний	человек	машина	человек, машина
Подход к выявлению основных компонентов системы	эксперт	автоматический	автоматический
Агенты	человеческие агенты	машинные агенты	человеческие и машинные агенты
Использование компетенций инженера знаний	да	нет	нет
Использование компетенций эксперта	да	нет	да
Системное представление предметной области	да	нет	да

ТАБЛИЦА 2. Инструменты предлагаемого метода валидации

TABLE 2. Tools for the Proposed Validation Method

Детерминанты качества метода валидации	Проблемы	Достигаемые результаты	Предлагаемые инструменты
Эксперты предметной области	субъективизм, неполное знание, ошибочность представления о системе	снижение субъективизма, выявление глубокого знания, выявление скрытого знания, изменение уровня собственного знания	верификация НКК, валидация НКК
Инженеры по знаниям	субъективизм, неполное знание	снижение субъективизма	алгоритм эффективных управлений
Схемы представления знаний	невозможность формализации в иерархическое представление	получение факторов системы, упорядоченных по значению	алгоритм эффективных управлений
Методы приобретения знаний	разработка методов выявления скрытого и глубокого знания	снижение субъективизма обнаружение знаний в результате алгоритмической обработки	алгоритм эффективных управлений
Проблемные области	специфика графовых моделей	системное представление предметной области, учет специфических свойств предметной области	верификация НКК, валидация НКК

Применение предложенного метода (см. рис. 1) позволяет осуществлять валидацию модели на системном уровне. Это означает возможность оценивать полученную НКК с точки зрения целевой направленности системы. Очевидно, что работа системы в целом, совокупная работа ее связей и

факторов, должны быть направлены на достижение результативности в первую очередь ее главных факторов, а не второстепенных или сопутствующих. Тогда критерием валидации становится смысловое содержание конструируемой модели системы в виде НКК, а не математическое содержание алгоритмов.

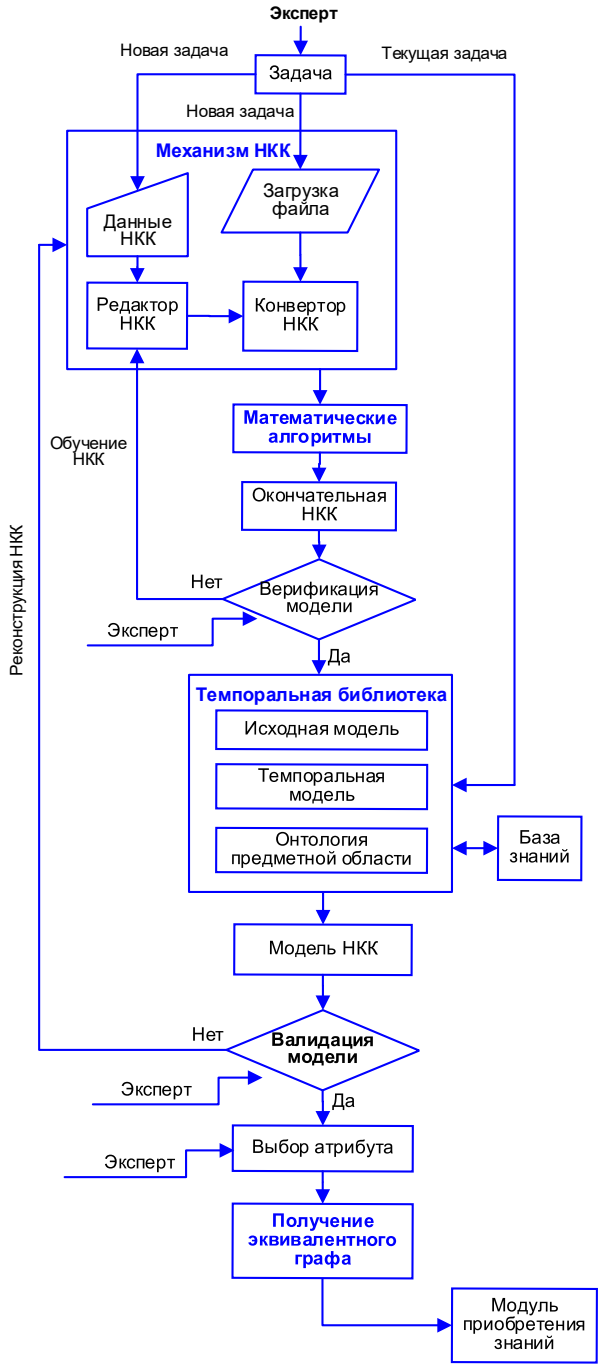


Рис. 1. Метод валидации графовых моделей

Fig. 1. Method for Validating Graph Models

Алгоритм эффективных управлений

Рассмотрим граф $G = \langle V, E \rangle$, где $V = \{V_1, V_2, \dots, V_N\}$ – множество вершин, а $E \in V \times V$ – множество дуг. Граф определяется матрицей смежности $\mathbf{A}^T = \|a_{i,j}\|_n$, где $a_{i,j}$ – вес дуги из вершины V_i в вершину V_j на интервале $-1 \le a_{i,j} \le 1$.

Рассматривается задача управления моделью роста с простейшим соотношением влияния и отклика:

$$(\boldsymbol{E} - \delta \boldsymbol{A}^T)\boldsymbol{x} = \boldsymbol{u},$$

где $\boldsymbol{E}$ – единичная матрица; $\boldsymbol{x} = (x_1\ x_2\ \dots\ x_n)^T$ – вектор воздействий, называемых откликом системы; $\boldsymbol{u} = (u_1\ u_2\ \dots\ u_n)^T$ – вектор управляющих воздействий (влияний); δ – коэффициент затухания $0 < \delta \le 1$.

Тогда задача максимизации отношения норм:

$$\frac{\|\boldsymbol{x}\|^2}{\|\boldsymbol{u}\|^2} = \frac{\boldsymbol{x}^T\boldsymbol{x}}{\boldsymbol{u}^T\boldsymbol{u}} \to \max,\ (\boldsymbol{V}^T\boldsymbol{V})^{-1}\boldsymbol{V}^T\boldsymbol{u} \ge \boldsymbol{0}_m,$$

где $(\boldsymbol{V}^T\boldsymbol{V})^{-1}\boldsymbol{V}^T\boldsymbol{u} \ge \boldsymbol{0}_m$ – условие неотрицательности коэффициентов разложения вектора $\boldsymbol{u}$ по векторам $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m$, решается относительно откликов при условии неотрицательности коэффициентов разложения вектора $\boldsymbol{x}$ по векторам $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_m$:

$$\frac{\boldsymbol{x}^T\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x}} \to \max,\ \boldsymbol{C}\boldsymbol{x} \ge \boldsymbol{0}_m, \text{ где}$$

$$\boldsymbol{C} = (\boldsymbol{V}^T\boldsymbol{V})^{-1}\boldsymbol{V}^T(\boldsymbol{E} - \delta\boldsymbol{A}^T),\ \boldsymbol{B} = (\boldsymbol{E} - \delta\boldsymbol{A})(\boldsymbol{E} - \delta\boldsymbol{A}^T).$$

Далее рассматривается задача квадратичного программирования:

$$\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x} \to \min,\ \boldsymbol{x}^T\boldsymbol{x} = \boldsymbol{1},\ \boldsymbol{C}\boldsymbol{x} \ge \boldsymbol{0}_m.$$

Функция Лагранжа для данной задачи имеет вид:

$$L = \boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x} + \boldsymbol{\lambda}(\boldsymbol{1} - \boldsymbol{x}^T\boldsymbol{x}) - \boldsymbol{\mu}^T\boldsymbol{C}\boldsymbol{x} \to \min.$$

Линеаризованная система необходимых условий минимума модифицированной функции Лагранжа по Берсекасу [10, 13] с регуляризацией по А.Н. Тихонову [11] имеет вид:

$$\begin{cases} \mathbf{H}\tilde{\mathbf{x}} = \tilde{\mathbf{G}}\tilde{\mathbf{y}} + 2(\alpha - \lambda)\boldsymbol{x} \\ \tilde{\mathbf{G}}^{\mathrm{T}}\tilde{\mathbf{x}} = \tilde{\mathbf{r}} \end{cases}, \text{ где}$$

$$\boldsymbol{H} = 2\bar{\mathbf{B}} - 2(\lambda - \alpha)\boldsymbol{E},\ \tilde{\mathbf{G}} = (2\boldsymbol{x} \vdots\ \tilde{\mathbf{C}}^T),$$
$$\tilde{\mathbf{r}}^T = (1 + \boldsymbol{x}^T\boldsymbol{x} \vdots\ \boldsymbol{0}_m^T),\ \boldsymbol{y}^T = (\hat{\lambda} \vdots\ \tilde{\boldsymbol{\mu}}^T).$$

На каждой итерации система может быть разрешена с помощью следующего алгоритма эффективных управлений с общими ограничениями:

1) решить систему линейных алгебраических уравнений (СЛАУ): $\tilde{\mathbf{T}} = \boldsymbol{H}^{-1}\tilde{\mathbf{G}}$ (k-ый столбец $\tilde{\mathbf{T}}$ является решением СЛАУ с матрицей $\boldsymbol{H}$ и правой частью, являющейся k-ым столбцом матрицы $\boldsymbol{G}$), $\boldsymbol{t} = 2(\alpha - \lambda)\boldsymbol{H}^{-1}\boldsymbol{x}$;

2) сформировать матрицу двойственной задачи $\tilde{\mathbf{D}} = \tilde{\mathbf{G}}^T\tilde{\mathbf{T}}$ и решить СЛАУ $\tilde{\mathbf{y}} = \tilde{\mathbf{D}}^{-1}(\tilde{\mathbf{r}} - \tilde{\mathbf{G}}^T\boldsymbol{t})$;

3) сформировать решение: $\tilde{\mathbf{x}} = \tilde{\mathbf{T}}\tilde{\mathbf{y}} + \boldsymbol{t}$;

4) нормализовать отклик: $\hat{\mathbf{x}} = \tilde{\mathbf{x}}/\|\tilde{\mathbf{x}}\|$, и выразить воздействие: $\hat{\mathbf{u}} = (\boldsymbol{E} - \delta\boldsymbol{A}^T)\hat{\mathbf{x}}$.

Детализированная версия алгоритма представлена в [14].

В результате работы алгоритма получаем две группы упорядоченных по убыванию факторов

системы (вершин графа) с соответствующими значениями их метрик: группы факторов отклика и факторов влияния. Важным достоинством представленного алгоритма является возможность настройки модели путем выбора объекта и вида управляющего воздействия. Это заключается в придании факторам модели следующих статусов: неуправляемые факторы, факторы, характеристики которых могут иметь только положительный рост или только отрицательный рост.

Вычислительный эксперимент

В качестве иллюстрации работы алгоритма был взят пример из [12]. В этой статье представлена НКК, разработанная тремя экспертами. В вычислительном эксперименте были рассчитаны компоненты векторов отклика и влияния для четырех НКК: три из них были разработаны экспертами, одна является результирующей, как показано в [12]. Результаты экспериментальных вычислений приведены в таблице 3.

ТАБЛИЦА 3. Результаты вычислительного эксперимента

TABLE 3. Results of the Computational Experiment

Номер вершины	Компонента вектора		Компонента вектора		Компонента вектора		Компонента вектора	
	отклика	влияния	отклика	влияния	отклика	влияния	отклика	влияния
	Модель 1 (Эксперт 1)		Модель 2 (Эксперт 2)		Модель 3 (Эксперт 3)		Модель 4 (результирующая НКК)	
1	0	0	0,1823	0,18231	–0,0464	–0,04638	–0,0164	–0,01644
2	0,0475	0,04752	0	0	0,1783	0,17828	0,1655	0,16548
3	0,0605	0,06046	0,3208	0,2619	0,267	0,25587	0,3354	0,30935
4	0,0438	0,03923	**0,6645**	**0,55822**	0,1415	0,10225	0,3163	0,27167
5	0,4546	0,45457	0,3197	0,31119	0,4153	0,34101	**0,4647**	**0,43138**
6	0,2609	0,22068	0,2828	0,27187	**0,4724**	**0,35747**	0,4454	0,39367
7	**0,6478**	**0,54785**	0,2292	0,2262	–0,2176	–0,20646	–0,0823	–0,08452
8	–0,1421	–0,14206	0,3083	0,30832	0,2584	0,17673	0,2037	0,18076
9	0,125	0,12218	0,1722	0,16851	0,2755	0,22957	0,2186	0,19913
10	–0,2343	–0,2343	–0,0434	–0,04342	0,2224	0,22545	0,2184	0,2186
11	0,193	0,19184	0,184	0,17477	0,2809	0,21288	0,3116	0,28528
12	0,2987	0,27453	0,1244	0,11596	0,3882	0,27401	0,323	0,27732
13	–0,284	–0,28396	0,1077	0,10769	–0,1	–0,09243	–0,0179	–0,01485

Примечание: цветом выделены вершины, получившие максимальные значения соответствующих компонент векторов, и, соответственно, занявшие первых два места; максимальные значения компонент векторов выделены дополнительно шрифтом.

Анализ полученных результатов показывает, что эксперты по-разному смотрят на систему в целом. Для НКК, разработанного экспертом 1, основным управляющим фактором является 7-ой фактор, для эксперта 2 – 4-ый фактор, для эксперта 3 – 6-ой (выделен овалом на рисунке 2).

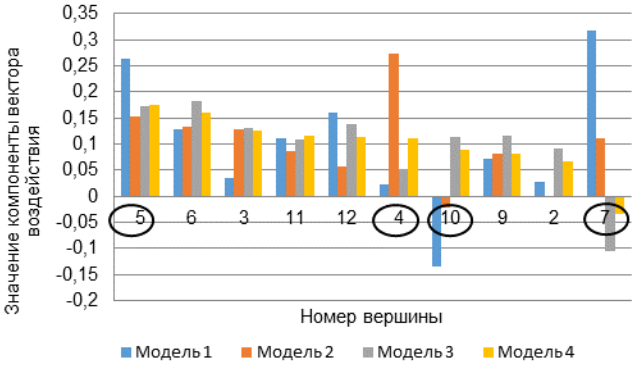


Рис. 2. Ранжирование факторов для четырех моделей

Fig. 2. Ranking Factors for Four Models

В результирующей НКК основной фактор – 5-ый. Различия в понимании системы сказались и на рангах всех остальных факторов, особенно по факторам 4, 5, 7 и 10. Так, 5-ый фактор не поставил на 1-ое место ни одного из экспертов. Однако в итоговой карте он оказался на первом месте. Как видно из приведенного примера, усреднение отношений может дать неверный результат, использование которого не даст ожидаемого эффекта для работы системы в целом. Эксперимент иллюстрирует различные взгляды экспертов на исследуемую проблему. Как было справедливо отмечено в [12], НКК «также показывает наличие неточных незнаний эксперта». В результате интеграции таких разных взглядов на проблему, получаемая результирующая НКК (модель 4) не отражает точку зрения ни одного из трех экспертов, а выражает некое совсем другое представление работы исследуемой системы с другими процессами, порождаемыми другим сочетанием факторов и причинно-следственных связей между ними. Кумулятивный индекс эффективности (для компонент вектора отклика системы) и кумулятивный индекс управляемости (для компонент вектора воздействий) для модели 1 и модели 4 (рисунок 3) подтверждают указанные различия. Для примера была взята небольшая модель,

состоящая всего из 13 вершин, однако даже в таком простом случае определить только из одношаговых связей важность факторов, выделить из них главные (например, первые два-три фактора) и представить себе действие системы в целом очень затруднительно. Если система состоит из большого числа взаимосвязанных факторов, определение главных факторов простыми методами представляет собой трудно решаемую задачу. Результаты эксперимента показывают необходимость пересмотра разрабатываемой модели исследуемой системы на предмет ее соответствия ментальным представлениям эксперта.

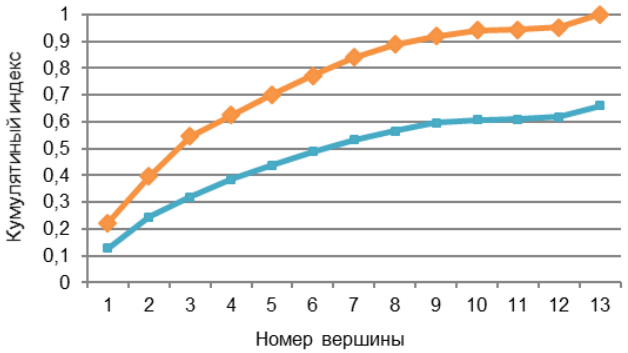


Рис. 3. Показатели кумулятивного индекса эффективности и управляемости для модели 1 (◆)и модели 4 (■)

Fig. 3. Indicators of the Cumulative Index of Efficiency and Manageability for Model 1 (◆) and for Model 4 (■)

Анализ результатов

Для оценки предложенной модели применяются следующие критерии.

Применимость для предметной области

Данный подход основан на теории систем и спектральных свойствах матричных представлений, ориентированных взвешенных знаковых графов, представляющих нечеткие когнитивные модели. Метод не накладывает ограничений на ориентированность ребер, знак и диапазон значений весов ребер. Используемые при решениях систем линейных уравнений матрицы будут в любых случаях положительными полуопределенными и симметричными.

Эффективность

Предложенный подход предоставляет возможность оценки достоверности и правильности построенной модели в текущий момент и по истечении релевантного периода. Такая оценка производится на основе сравнения полученных результатов с ментальными представлениями экспертов о работе системы. В случае неудовлетворительной оценки, происходит доработка НКК (дополнение или изъятие вершин и/или связей, изменение весов связей и т. п.). После корректировки вновь производится расчет метрик вершин и сравнение. Этот процесс повторяется до достижения приемлемых результатов.

Новизна полученных результатов

Получаемые результаты являются новыми обнаруженными знаниями, скрытыми в самой НКК. Эти знания об исследуемой системе являются ранее неизвестными для экспертов, разрабатывающих НКК. Вычисленные метрики вершин выражают направленность системы. Под направленностью системы мы понимаем проекцию собственного вектора, соответствующего минимальному собственному числу матрицы квадратичной форы (B) на множество допустимых решений, определяемое набором ограничений.

Качество полученных результатов

Результат работы алгоритма основан на спектральных свойствах матриц и не предусматривает эвристических алгоритмов. Решение задачи оптимизации на каждой итерации в исходных неизвестных обеспечивает точность решения на исходных уравнениях.

Сложность полученного решения

Вычислительная сложность алгоритма эффективных управлений, используемого в предложенном методе валидации графовых моделей, составляет $O(n^3)$ арифметических операций. Время решения алгоритма полиномиально и составляет доли секунды. Подбор значения параметра регуляризации требует дополнительных временны́х затрат.

Формат результатов

Достижение приемлемого уровня информативности получаемых результатов осуществляется путем их представления в виде упорядоченных по убыванию метрик вершин графовой модели для эксперта. Процесс сравнения является максимально простым. Сравнению подвергаются только несколько ключевых вершин.

Реализуемость

Матрицы задачи составляют матрицы вторых производных квадратичных форм, т. е. являются симметричными и знакоопределенными по построению. Обусловленность возникающих СЛАУ улучшается регуляризацией по Тихонову [11]. Используемый метод множителей Лагранжа требует определения только тех двойственных переменных, которые соответствуют активным ограничениям.

Заключение

В этой статье мы рассматриваем проблему валидации графовых моделей сложных систем. Основным вкладом данной работы является предложение нового метода валидации графовых моделей на основе алгоритма эффективных управлений. Валидация НКК производится по критерию соответствия смысловому содержанию НКК. Это соответствие достигается сравнением упорядоченного по убыванию набора главных вершин, получивших наибольшие показатели эффективности, с набором вершин, которые являются глав-

ными согласно ментальным представлениям эксперта. Используется разработанный нами алгоритм эффективных управлений, определяющий направленность собственного вектора матрицы смежности системы. Это свойство определяет критерий валидации сконструированной НКК.

В результате мы получаем новые знания о работе системы в целом, которые и будут представлять собой неявные, скрытые знания, добытые из нечеткой когнитивной модели предметной области. В этом случае НКК представляет собой «первичные знания» о системе, полученные от экспертов. Качество получаемого результата при использовании НКК рассматривается с точки зрения возможностей верификации и валидации разрабатываемой модели исследуемой системы.

БЛАГОДАРНОСТИ

Работа выполнена при поддержке гранта Российского фонда фундаментальных исследований № 19-01-00109.

Список используемых источников

1. Kosko B. Fuzzy cognitive maps // International Journal of Man-Machine Studies. 1986. Vol. 24. Iss. 1. PP. 65–75. DOI:10.1016/S0020-7373(86)80040-2
2. Salmeron J.L., Mansouri T., Moghadam M.R.S., Mardani A. Learning Fuzzy Cognitive Maps with modified asexual reproduction optimisation algorithm // Knowledge-Based Systems. 2019. Vol. 163. PP. 723–735. DOI:10.1016/j.knosys.2018.09.034
3. Konar A., Chakraborty U.K. Reasoning and unsupervised learning in a fuzzy cognitive map // Information Sciences. 2005. Vol. 170. Iss. 2-4. PP. 419–441. DOI:10.1016/j.ins.2004.03.012
4. Hebb D.O. The Organization of Behavior: A Neuropsychological Theory. London: Psychology Press, 2005. 335 p.
5. Kumbasar T., Eksin İ., Güzelkaya M., Yeşil E. Big Bang Big Crunch Optimization Method Based Fuzzy Model Inversion // Proceedings of the 7th Mexican International Conference on Artificial Intelligence on Advances in Artificial Intelligence (MICAI 2008, Atizapán de Zaragoza, Mexico, 27–31 October 2008). Lecture Notes in Computer Science. Vol. 5317. Berlin, Heidelberg: Springer, 2008. PP. 732–740. DOI:10.1007/978-3-540-88636-5_69
6. Vascak J. Approaches in adaptation of fuzzy cognitive maps for navigation purposes // Proceedings of the 8th International Symposium on Applied Machine Intelligence and Informatics (SAMI, Herlany, Slovakia, 28–30 January 2010). IEEE, 2010. PP. 31–36. DOI:10.1109/SAMI.2010.5423716
7. Papageorgiou E., Stylios C., Groumpos P. Fuzzy Cognitive Map Learning Based on Nonlinear Hebbian Rule // Proceedings of the 16th Australian Conference on Advances in Artificial Intelligence (AI, Perth, Australia, 3–5 December 2003). Lecture Notes in Computer Science. Vol. 2903. Berlin: Springer, Heidelberg, 2003. PP. 256–268. DOI:10.1007/978-3-540-24581-0_22
8. Leu G., Abbass H. A multi-disciplinary review of knowledge acquisition methods: From human to autonomous eliciting agents // Knowledge-Based Systems. 2016. Vol. 105. PP. 1–22. DOI:10.1016/j.knosys.2016.02.012
9. Tselykh A.N., Vasilev V., Tselykh L., Barkovskii S.A. Method Maximizing the Spread of Influence in Directed Signed Weighted Graphs // Advances in Electrical and Electronic Engineering. 2017. Vol. 15. Iss. 2. DOI:10.15598/aeee.v15i2.1950
10. Bertsekas D.P. Constrained Optimization and Lagrange Multiplier Methods. Belmont: Athena Scientific, 1996. PP. 158–297.
11. Tikhonov A., Arsenin V. Solutions of Ill-Posed Problems. New York: Wiley, 1977. 272 p.
12. Banini G.A., Bearman R.A. Application of fuzzy cognitive maps to factors affecting slurry rheology // International Journal of Mineral Processing. 1998. Vol. 52. Iss. 4. PP. 233–244. DOI:10.1016/S0301-7516(97)00071-9
13. Bertsekas D.P. The Method of Multipliers for Equality Constrained Problems // Constrained Optimization and Lagrange Multiplier Methods. New York: Elsevier, 1982. PP. 95–157.
14. Tselykh A., Vasilev V., Tselykh L. Ferreira F.A.F. Influence control method on directed weighted signed graphs with deterministic causality // Annals of Operations Research. 2020. DOI:10.1007/s10479-020-03587-8
15. Tselykh A., Vasilev V., Tselykh L. Assessment of influence productivity in cognitive models // Artificial Intelligence Review. 2020. DOI:10.1007/s10462-020-09823-8

* * *

Method for Validating Graph Models Based on the Effective Control Algorithm

V. Vasilev[1], A. Tselykh[1], L. Tselykh[2]

[1]Institute of Computer Technologies and Information Safety of Southern Federal University, Taganrog, 347900, Russian Federation
[2]Taganrog Institute named after A.P. Chekhov (branch) of Rostov State University of Economics, Taganrog, 347936, Russian Federation

Article info
DOI:10.31854/1813-324X-2020-6-3-58-65
Received 1st August 2020
Accepted 19th August 2020

For citation: Vasiliev V., Tselykh A., Tselykh L. Method for Validating Graph Models Based on the Effective Control Algorithm. *Proc. of Telecom. Universities.* 2020;6(3):58–65. (in Russ.) DOI:10.31854/1813-324X-2020-6-3-58-65

Abstract: *The article proposes a method for validating mathematical models represented by oriented weighted signed graphs using an efficient control algorithm. The method considers the validated model in terms of spectral properties of the graph adjacency matrix represented by a fuzzy cognitive map (FCM). Using an efficient control algorithm, you can determine the eigenvector direction of the adjacency matrix. This property defines the criteria for checking the FCM.*

Keywords: *model validation, directed weighted graph, fuzzy cognitive map, efficient management.*

References

1. Kosko B. Fuzzy cognitive maps. *International Journal of Man-Machine Studies*. 1986;24(1):65–75. DOI:10.1016/S0020-7373(86)80040-2
2. Salmeron J.L., Mansouri T., Moghadam M.R.S., Mardani A. Learning Fuzzy Cognitive Maps with modified asexual reproduction optimisation algorithm. *Knowledge-Based Systems*. 2019;163:723–735. DOI:10.1016/j.knosys.2018.09.034
3. Konar A., Chakraborty U.K. Reasoning and unsupervised learning in a fuzzy cognitive map. *Information Sciences*. 2005; 170(2-4):419–441. DOI:10.1016/j.ins.2004.03.012
4. Hebb D.O. *The Organization of Behavior: A Neuropsychological Theory*. London: Psychology Press; 2005. 335 p.
5. Kumbasar T., Eksin İ., Güzelkaya M., Yeşil E. Big Bang Big Crunch Optimization Method Based Fuzzy Model Inversion. *Proceedings of the 7th Mexican International Conference on Artificial Intelligence on Advances in Artificial Intelligence, MICAI 2008, 27–31 October 2008, Atizapán de Zaragoza, Mexico. Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer; 2008. vol.5317. p.732–740. DOI:10.1007/978-3-540-88636-5_69
6. Vascak J. Approaches in adaptation of fuzzy cognitive maps for navigation purposes. *Proceedings of the 8th International Symposium on Applied Machine Intelligence and Informatics, SAMI, 28–30 January 2010, Herlany, Slovakia*. IEEE; 2010. p.31–36. DOI:10.1109/SAMI.2010.5423716
7. Papageorgiou E., Stylios C., Groumpos P. Fuzzy Cognitive Map Learning Based on Nonlinear Hebbian Rule. *Proceedings of the 16th Australian Conference on Advances in Artificial Intelligence, AI, 3–5 December 2003, Perth, Australia. Lecture Notes in Computer Science*. Berlin: Springer, Heidelberg; 2003. vol.2903. p.256–268. DOI:10.1007/978-3-540-24581-0_22
8. Leu G., Abbass H. A multi-disciplinary review of knowledge acquisition methods: From human to autonomous eliciting agents. *Knowledge-Based Systems*. 2016;105:1–22. DOI:10.1016/j.knosys.2016.02.012
9. Tselykh A.N., Vasilev V., Tselykh L., Barkovskii S.A. Method Maximizing the Spread of Influence in Directed Signed Weighted Graphs. *Advances in Electrical and Electronic Engineering*. 2017;15(2). DOI:10.15598/aeee.v15i2.1950
10. Bertsekas D.P. *Constrained Optimization and Lagrange Multiplier Methods*. Belmont: Athena Scientific; 1996. p.158–297.
11. Tikhonov A., Arsenin V. Solutions of Ill-Posed Problems. New York: Wiley; 1977. 272 p.
12. Banini G.A., Bearman R.A. Application of fuzzy cognitive maps to factors affecting slurry rheology. *International Journal of Mineral Processing*. 1998;52(4):233–244. DOI:10.1016/S0301-7516(97)00071-9
13. Bertsekas D.P. The Method of Multipliers for Equality Constrained Problems // Constrained Optimization and Lagrange Multiplier Methods. New York: Elsevier, 1982. p.95–157.
14. Tselykh A., Vasilev V., Tselykh L. Ferreira F.A.F. Influence control method on directed weighted signed graphs with deterministic causality. *Annals of Operations Research*. 2020. DOI:10.1007/s10479-020-03587-8
15. Tselykh A., Vasilev V., Tselykh, L. Assessment of influence productivity in cognitive models. *Artificial Intelligence Review*. 2020. DOI:10.1007/s10462-020-09823-8

Сведения об авторах:

ВАСИЛЬЕВ Владислав Сергеевич — кандидат физико-математических наук, доцент кафедры информационно-аналитических систем безопасности им. профессора Берштейна Леонида Самойловича Института компьютерных технологий и информационной безопасности Южного федерального университета, vsvasilev@sfedu.ru
https://orcid.org/0000-0001-7485-8614

ЦЕЛЫХ Александр Николаевич — доктор технических наук, профессор, заведующий кафедрой информационно-аналитических систем безопасности им. профессора Берштейна Леонида Самойловича Института компьютерных технологий и информационной безопасности Южного федерального университета, ant@sfedu.ru
https://orcid.org/0000-0001-6956-5315

ЦЕЛЫХ Лариса Анатольевна — кандидат экономических наук, доцент кафедры экономики и предпринимательства Таганрогского института им. А.П. Чехова (филиала) Ростовского государственного экономического университета, l.tselykh58@gmail.com
https://orcid.org/0000-0001-5663-1563

УДК 004.056.5(076) DOI:10.31854/1813-324X-2020-6-3-66-74

О сравнении систем защиты информации при асимптотическом управлении информационной безопасностью КИИ

С.Д. Ерохин[1], **А.Н. Петухов**[1, 2*], **П.Л. Пилюгин**[1, 3]

[1]Московский технический университет связи и информатики,
Москва, 111024, Российская Федерация
[2]Московский институт электронной техники,
Москва, 124498, Российская Федерация
[3]Московский государственный университет им. М.В. Ломоносова
Москва, 119991, Российская Федерация
[*]Адрес для переписки: anpetukhov@yandex.ru

Информация о статье
Поступила в редакцию 19.08.2020
Принята к публикации 08.09.2020

Ссылка для цитирования: Ерохин С.Д., Петухов А.Н., Пилюгин П.Л. О сравнении систем защиты информации при асимптотическом управлении информационной безопасностью КИИ // Труды учебных заведений связи. 2020. Т. 6. № 3. С. 66–74. DOI:10.31854/1813-324X-2020-6-3-66-74

Аннотация: *В статье рассматриваются возможности управления безопасностью критических информационных инфраструктур. Предлагаются подходы к построению политики, неориентированной на оценку остаточного риска и фиксированный список угроз. Обосновывается возможность построения управления информационной безопасностью на основе мониторинга событий безопасности. Предлагается формальное описание событий безопасности, механизмов защиты. Вводится отношение порядка для сравнения систем защиты информации и осуществления асимптотического управления информационной безопасностью критических информационных инфраструктур.*

Ключевые слова: *информационная безопасность, политика безопасности, события безопасности, мониторинг событий безопасности, асимптотическое управление, характеристики сетевых событий, критические информационные инфраструктуры.*

Введение

Защищенность в традиционном смысле характеризует эффективность усилий, направленных на противостояние некоторому установленному потенциалу агрессивности среды и выражает достаточность реагирования на возможные воздействия, нарушающее нормальное функционирование защищаемого объекта. Метризация защищенности в рамках рискориентированного подхода строится на основе модели с полным перекрытием Клементса [1], которая рассматривает степень защищенности как априорную характеристику отношений угрозы, актива и защитной меры, а высокоуровневые сущности «Общих критериев» [2, 3, 4] включают параметр «риска» как интегрального остатка от неполной защищенности.

В идеале спецификация агрессии должна включать описание всей причинно-следственной цепочки проявления вредоносного воздействия, а именно «направление вредоносного намерения или потенциала (угроза) – организационно-технический дефект или несовершенство (уязвимость) – последовательность шагов по реализации (атака) – характер ущерба от снижения или утраты информационными активами своей ценности или функциональности (инцидент)». Если предположить гипотетическую возможность исчерпывающего (предельно детального) такого описания, не допускающего вариативности и интерпретации в своих частях, то можно попытаться решить задачу противопоставления специфицированной таким образом агрессии некоторой защитной мерой, обеспечивающей полную защищенность от этой агрессии.

Однако гарантированно исчерпывающе перечислить такие «точечные» описания пока не представляется возможным, поэтому для спецификации агрессии используются обобщающие категории, включающие возможность вариативной интерпретации (в основном, в реализационной части). Таким образом, спецификация вида агрессии объединяет множество различных интерпретаций причинно-следственных цепочек, и нет никаких оснований

утверждать (несмотря на заверения производителя), что противопоставляемые ей защитные меры охватывают и прерывают все интерпретации. Чем больше доля такого охвата, тем выше уровень защищенности. Это полностью соответствует определению защищенности как способности противостоять агрессивному вредоносному воздействию на защищаемый объект.

Для оценки защищенности предлагается использовать вероятность возникновения угрозы и вероятность ее реализации (как правило, без спецификации сценария атаки и других деталей). Эти характеристики носят исключительно экспертный характер (кроме характеристик, связанных с криптостойкостью), и они позволяют рассматривать защищенность как характеристику обратную остаточному риску.

Подход на основе асимптотического управления информационной безопасностью [5] критических информационных инфраструктур (КИИ) предполагает отказ от оценки допустимого (остаточного) риска для измерения эффективности применяемых средств защиты. В связи с этим возникает естественный вопрос о соотношении (хотя бы качественном) достигаемого уровня безопасности при использовании тех или иных защитных мер. При этом для любого суждения о *степени* защищенности (оценка или сопоставление) всегда постулируется «принцип монотонности»: защищенность объекта при совокупном использовании двух любых защитных мер не ниже защищенности при использовании любой из них в отдельности.

В процедуре оценки рисков информационной безопасности ISO 27005 рассматриваются два подхода к оценке риска: качественный и количественный. «Как правило, вначале применяют качественный анализ для выделения высокоприоритетных рисков, а затем уже для выявленных рисков применяют количественный анализ, который является более трудоемким и дает более точные результаты» [6, 7].

Для КИИ асимптотический подход к управлению информационной безопасности (ИБ) использует последовательное улучшение качества системы защиты. Поскольку оценка качества, базирующаяся на измерении в количественных шкалах остаточного риска, в рамках рассматриваемого подхода не применяется, следует использовать качественные (порядковые) шкалы для определения улучшения (направления развития) свойств систем защиты, а это, в свою очередь, требует ввести отношение порядка на множестве всевозможных систем защиты.

Традиционные количественные оценки ущерба также могут быть использованы для определения отношения порядка на множестве всевозможных систем защиты в рассматриваемом асимптотическом подходе к управлению ИБ КИИ. В настоящей работе для устранения описанных выше недостатков традиционной оценки ущерба предлагается альтернативный подход к сравнению качества защиты, основанный на субъектно-объектной модели системы защиты.

В качестве универсальной модели механизма защиты для формализации описания в асимптотическом управлении ИБ используется монитор событий безопасности [8]. В этом случае система защиты рассматривается, как совокупность механизмов защиты из множества $\boldsymbol{M}$, где механизм защиты выполняет две основных функции: обнаружение и противодействие. Для эффективности механизма необходимым является обнаружение, так как необнаруженным событиям нельзя противодействовать. Это позволяет рассматривать множество доступных для наблюдения монитором событий безопасности (из всего пространства таких событий), как основную характеристику механизма защиты [8].

Свойства инцидентов и событий безопасности, используемые в рассматриваемой модели

Событие безопасности в соответствии с ГОСТ РИСО/МЭК ТО 18044-2007 рассматривается как обнаруживаемое монитором (через набор признаков) состояние наблюдаемой системы, которое может привести к возникновению инцидента или просто индицирует возможность его возникновения. В обоих случаях формально это описывалось через корреляционную связь события с инцидентом: «вероятности возникновения события при условии инцидента $\eta = P$ (событие/инцидент есть) и при его отсутствии $\mu = P$ (события/инцидента нет) различны» [9]. Далее будем обозначать $\boldsymbol{A} = \{a_i\}$ множество событий безопасности, обнаруживаемых соответствующими механизмами безопасности из $\boldsymbol{M}$ конкретной системы защиты. Для дальнейшего изложения следует учитывать, что неважно, есть или нет точной оценки этих вероятностей, важно, только, как отмечено выше, чтобы они не были равны, иначе событие не будет информативным.

При этом важно, чтобы события безопасности были независимыми или «почти» независимыми, т. е. слабо корреляционно связаны между собой. Действительно нет никакого смысла наблюдать два события безопасности, если они возникают всегда совместно. Таким образом, события безопасности из группы событий, индицирующих конкретный инцидент, будем считать независимыми или «почти» независимыми. Тогда, исходя из правила добавления только независимых событий, можно считать, что будут независимы и все события из $\boldsymbol{A}$.

Кроме того, события из $\boldsymbol{A}$ будут совместными, так как возможно наблюдение нескольких независимых признаков одновременно или в период наблюдения.

Вместе с тем следует иметь в виду, что событие может индицировать несколько инцидентов. Однако значимость (ценность) такого события для

индикации тем меньше, чем большее число инцидентов им индицируется. Это объясняется тем, что факт наблюдения такого события не дает возможность точно индицировать конкретный инцидент и, соответственно, запускать процедуру конкретного реагирования. Такое событие может использоваться, как правило, для подтверждения опасности и повышения уверенности при наблюдении других событий безопасности. В принципе возможен вариант, когда несколько таких «неоднозначных» событий указывают на один конкретный инцидент (множества индицируемых ими инцидентов в качестве пересечения имеют один элемент). Однако при дальнейшем рассмотрении, будем предполагать (предположение П-1), что таких «неоднозначных» событий безопасности нет.

Под инцидентом $\boldsymbol{I_k}$ будем понимать событие, связанное с нарушением некоторого критического процесса системы, приводящее к прекращению (полному или частичному) функционирования КИИ. В отношении таких инцидентов предполагается:

– во-первых, что сам инцидент также наблюдаем средствами мониторинга (что далеко не всегда возможно, но допустимо для мониторинга ИТКС и сетей электросвязи);

– во-вторых, инцидент в данном случае рассматривается как критическое нарушение функционирования КИИ (или нарушение критического процесса КИИ [10]), это позволяет уточнить общее определение инцидента в ГОСТ РИСО/МЭК ТО 18044-2007, как «нежелательного или неожиданного события(й) ИБ, с которыми связана значительная вероятность компрометации бизнес-операций и создания угрозы ИБ [11]»;

– в-третьих, инциденты рассматриваются, как независимые или «почти» независимые события (т. е. слабо корреляционно связанные), так как даже, когда они наступают совместно, структурно они могут соответствовать разным компонентам системы; можно предположить, что в этом случае у них есть общая причина, которая и должна рассматриваться как инцидент;

– в-четвертых, предполагается, что состав инцидентов фиксирован (предположение П-2): для рассматриваемой задачи управления ИБ предполагается, что все критические процессы (возможные инциденты) были идентифицированы ранее в соответствии с инвентаризацией КИИ при категорировании [10]; изменение состава инцидентов возможно на основе получения дополнительной информации, и новому составу инцидентов будет соответствовать новая система защиты информации (СЗИ), улучшать свойства которой можно на основе мониторинга в процессе асимптотического управления.

Последнее обстоятельство подчеркивает кусочно-непрерывный характер процесса асимптотического управления. Текущее функционирование адаптивной составляющей в процессе «непрерывного» периода асимптотического управления накапливает некоторую «критическую массу» оценки прогностической составляющей и может инициировать качественный скачок, связанный с изменением состава инцидентов. Кроме того, изменение номенклатуры (состава и описания) инцидентов может осуществляться извне, например, по каналам взаимодействия с ГосСОПКА.

Сравнение систем защиты информации на основе наблюдаемых событий безопасности

Сравнение СЗИ строится на основе модели системы защиты с полным перекрытием. Однако в предлагаемом подходе происходит ее трансформация этой модели от структуры «угроза – механизм защиты – объект защиты» к структуре «угроза – событие безопасности – механизм защиты – инцидент безопасности – объект защиты». Т. е. вместо тройки (t_i, m_j, o_k) рассматривается тройка (a_i, m_j, I_k), где $\{t_i\}$ – угрозы; $\{m_j\}$ – механизмы; $\{o_k\}$ – объекты; $\{a_i\}$ – события; $\{I_k\}$ – инциденты (рисунок 1).

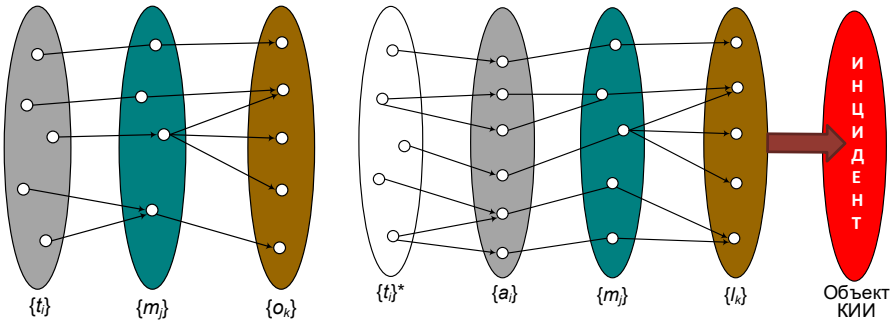


Рис. 1. Трансформация модели с полным перекрытием

Fig. 1. Transformation of the Model with Full Overlap

Для сетевых механизмов защиты (МЭ, IDS, IPS, SIEM и т. д.) в процессе мониторинга индицируются события на основе сигнатур и обнаружения аномалий [12]. Отметим, что понятие сигнатуры событий безопасности рассматривается здесь несколько шире, чем аналогичное понятие, например, в практике антивирусной деятельности (фрагмент программного кода, характерный для вредоносного

ПО). Сигнатура события безопасности (шаблон, паттерн, комплекс признаков и др.) предназначена для обнаружения факта, свидетельствующего, что соответствующее событие или группа событий произошли. Сигнатура может определяться источниками информации о событии (например, сенсоры первичных измерений) и структурой формирования из этой информации более сложных агрегатов, которые соответствуют комплексным событиям, коррелированным с инцидентом безопасности (схема агрегации).

В первом случае можно рассматривать набор (вектор) из n параметров последовательных состояний системы $s_i = (p_1, ... p_n)$, как представление события безопасности (или набора таких событий), описываемых сигнатурой некоторого инцидента I_m. Здесь $(p_1, ... p_n)$, это наборы параметров состояний системы наблюдаемых в различные моменты времени интервала наблюдений, что соответствует приведенным выше определениям события безопасности и инцидента.

Таким образом, мы можем далее рассматривать соответствие наблюдаемых параметров системы сигнатуре как выявление (наблюдение) соответствующих событий безопасности. Т. е. если $\boldsymbol{S} = \{s_i\}$ множество всех возможных сигнатур, то в соответствии с этим уточнением положим $\theta(s_i) \subseteq \boldsymbol{A}$, где θ: $\boldsymbol{S} \rightarrow \boldsymbol{A}$ отображение множества сигнатур на множество событий безопасности. Это отображение может быть не взаимно-однозначным, в этом случае θ^{-1} может рассматриваться как метод агрегирования событий безопасности $\{\theta^{-1}(a_1, ... a_r)\} = \{s_i\}$, а $\{s_i / \theta(s_i) \subseteq \boldsymbol{A}\} \subseteq \boldsymbol{S}$ – как новый набор независимых событий безопасности, описываемых сигнатурами.

Во втором случае монитор позволяет обнаруживать некоторые отклонения от ожидаемого «поведения» системы. При этом следует иметь в виду, что аномалии сетевого трафика могут являться вполне безопасным явлением [13]. Поэтому идентификация аномалий как событий безопасности, соответствующих инцидентам, требует дополнительной фильтрации и обработки получаемых данных [14], причем для этого привлекаются не только вероятностные методы [15]. Такие аномалии, идентифицированные как события безопасности, также отслеживаются монитором на основе соответствующих сигнатур (например, превышение попыток ввода пароля или одновременное открытие большого числа TCP-соединений и т. п.).

Так как далее мониторинг на основе сигнатурного обнаружения событий полагается общей моделью используемых механизмов защиты, то в дальнейшем будем исходить из того, что $\boldsymbol{M} = \{s_i / \theta(s_i) \subseteq \boldsymbol{A}\}$, т. е. все механизмы защиты описываются множеством наблюдаемых ими событий безопасности, что соответствует множеству сигнатур инцидентов. Используя эти утверждения, покажем возможность сравнивать системы защиты, использующие мониторинг на основе сигнатурного обнаружения событий.

Основой для сравнительной оценки эффективности различных СЗИ могут быть множества используемых ими сигнатур событий безопасности $\boldsymbol{S}_k \subseteq \boldsymbol{S}$ (эксплуатационные и архитектурные параметры здесь не рассматриваются). Для того, чтобы сравнить СЗИ, можем сопоставить множества всех событий безопасности СЗИ, устанавливая тем самым на множестве СЗИ отношение частичного порядка. Исходя из этого, далее будут рассмотрены условия введения отношения порядка $\succcurlyeq_S$ на основе сравнения множеств сигнатур СЗИ для группы событий безопасности.

Отношение частичного порядка на множествах сигнатур событий безопасности

Пусть множество инцидентов содержит только один инцидент (предположение П-3) и пусть, как это описано выше, $СЗИ_1$ и $СЗИ_2$ полностью характеризуются их множествами сигнатур: S_1 и S_2, тогда при сравнении множеств сигнатур возможны следующие случаи:

1) $\boldsymbol{S}_1 \supseteq \boldsymbol{S}_2$ или $\boldsymbol{S}_1 \subseteq \boldsymbol{S}_2$ – отношения включения множеств устанавливает для СЗИ (и соответствующих им множеств событий безопасности) отношение предпочтения: $\boldsymbol{S}_1 \succcurlyeq_S \boldsymbol{S}_2$ или $\boldsymbol{S}_1 \preccurlyeq_S \boldsymbol{S}_2$;

2) $\boldsymbol{S}_1 \cap \boldsymbol{S}_2 = \emptyset$ – сравнение на основе включения множеств не применимо;

3) $\boldsymbol{S}_1 \cap \boldsymbol{S}_2 \neq \emptyset$, но $\boldsymbol{S}_1 \not\supseteq \boldsymbol{S}_2$ или $\boldsymbol{S}_1 \not\subseteq \boldsymbol{S}_2$ – сравнение на основе включения множеств также не применимо.

Для повышения безопасности КИИ на основе асимптотического управления [16] первый случай показывает, как последовательное «улучшение» эффективности защиты связано с добавлением новых сигнатур в системы мониторинга. Однако возможно возникновение различных ситуаций, соответствующих случаю 2. Например, при агрегировании нескольких сигнатур в одну (построение сигнатуры-шаблона) или при выявлении (и удалении) «неэффективных» или менее эффективных сигнатур и замены их другими, что особенно важно в случае ограниченности ресурсов. В этом случае рассматриваются не множества сигнатур, а соответствующие им события безопасности (т. е. сигнатура рассматривается как результат агрегирования событий безопасности). Далее сравнение производится, если это возможно, как сравнение множеств. Например, пусть $\boldsymbol{A}_i$ множество событий безопасности, соответствующее множеству сигнатур $\boldsymbol{S}_i$: $\boldsymbol{A}_i = \theta(\boldsymbol{S}_i)$. Если справедливо, что: $\boldsymbol{A}_1 \supset \boldsymbol{A}_2$ или $\boldsymbol{A}_1 \subset \subset \boldsymbol{A}_2$, тогда отношения включения множеств устанавливает для СЗИ (и соответствующих им множеств сигнатур и событий безопасности) отношение предпочтения $\boldsymbol{S}_1 \succcurlyeq_S \boldsymbol{S}_2$ или $\boldsymbol{S}_1 \preccurlyeq_S \boldsymbol{S}_2$.

В других ситуациях возможны различные варианты проведения сравнений, которые не столь очевидны. Если $\boldsymbol{S}_1 \cap \boldsymbol{S}_2 = \emptyset$, то для исключения возможных неточностей также рассматриваем множества соответствующих им событий безопасности. В этом случае также имеем, что $\boldsymbol{A}_1 \cap \boldsymbol{A}_2 = \emptyset$, и для проведения «грубой оценки» множеств сигнатур в качестве метрики множества событий безопасности используем мощность соответствующего множества $\|\boldsymbol{A}_i\| = |\boldsymbol{A}_i|$.

Такой линейной метрике соответствует сложение вероятностей событий безопасности, которое здесь не применимо, так как события хотя и независимы, но совместны. Однако в качестве обоснования такой на первый взгляд тривиальной метрики рассмотрим вероятность $P_I(\boldsymbol{A})$ индицирования инцидента группой $\boldsymbol{A}$ соответствующих ему совместных событий безопасности. Будем предполагать, что все события безопасности одинаково значимы для обнаружения инцидента (предположение П-4), и что вероятности таких событий примерно совпадают. Инцидент индицируется, если происходит обнаружение хотя бы одного события безопасности и, соответственно, $P_I(\boldsymbol{A}) = 1 - P(\bar{\boldsymbol{A}})$, где $P(\bar{\boldsymbol{A}})$ – вероятность того, что не произойдет ни одного события безопасности из $\boldsymbol{A} = (a_1, a_2, \ldots, a_n)$. Тогда $\bar{\boldsymbol{A}} == (\bar{a}_1 \cup \bar{a}_2 \cup \bar{a}_2 \ldots \cup \bar{a}_n)$ – произведение независимых событий и $P(\bar{\boldsymbol{A}}) = \prod_1^n (1 - p(a_i)) = (1-p)^n$, где $p \approx p(a_i)$ вероятности событий a_i, если они равны (или это среднее значение этих вероятностей). Отсюда видно, что $P_I(\boldsymbol{A}) = 1 - (1-p)^n$ при росте n монотонно возрастает.

Очевидно, что при больших количествах событий безопасности n изменения $P_I(\boldsymbol{A})$ будут невелики, и при традиционной оценке остаточных рисков ими можно пренебречь, однако в случае КИИ любое улучшение будет значимым. При этом отметим, что возможно уменьшение показателя степени в функции $P_I(\boldsymbol{A}) = 1 - (1-p)^{n-r}$, где r обозначает минимальное число событий безопасности необходимое для обеспечения нужного качества индикации инцидента в соответствии со схемой последовательных испытаний Бернулли (рисунок 2).

Для определения $\|\boldsymbol{A}_1\| \geqq \|\boldsymbol{A}_2\|$ или $\|\boldsymbol{A}_1\| \leqq \|\boldsymbol{A}_2\|$ можно просто провести сравнение по количеству событий безопасности и, соответственно, получим: $\boldsymbol{S}_1 \succcurlyeq_S \boldsymbol{S}_2$ или $\boldsymbol{S}_1 \preccurlyeq_S \boldsymbol{S}_2$. В случае, если $\boldsymbol{A}_1 \cap \boldsymbol{A}_2 \neq \emptyset$, можно проводить сравнение множеств $\boldsymbol{B}_1 = \boldsymbol{A}_1 \setminus (\boldsymbol{A}_1 \cap \boldsymbol{A}_2)$ и $\boldsymbol{B}_2 = \boldsymbol{A}_2 \setminus (\boldsymbol{A}_1 \cap \boldsymbol{A}_2)$, тогда $\boldsymbol{B}_1 \cap \boldsymbol{B}_2 = \emptyset$ и $\|\boldsymbol{B}_1\| \geqq \|\boldsymbol{B}_2\|$ или $\|\boldsymbol{B}_1\| \leqq \|\boldsymbol{B}_2\|$, и, соответственно, $\boldsymbol{S}_1 \succcurlyeq_S \boldsymbol{S}_2$ или $\boldsymbol{S}_1 \preccurlyeq_S \boldsymbol{S}_2$ (рисунок 3).

С учетом установленных предположений и ограничений для введенного отношения предпочтения $\succcurlyeq_S$ можно утверждать, что:

$$\text{СЗИ}_1 \succcurlyeq_S \text{СЗИ}_2, \text{ если } \boldsymbol{S}_1 \succcurlyeq_S \boldsymbol{S}_2;$$

$$\text{СЗИ}_1 \preccurlyeq_S \text{СЗИ}_2, \text{ если } \boldsymbol{S}_1 \preccurlyeq_S \boldsymbol{S}_2;$$

$$\text{СЗИ}_1 =_S \text{СЗИ}_2, \text{ если } \boldsymbol{S}_1 \succcurlyeq_S \boldsymbol{S}_2 \text{ и } \boldsymbol{S}_1 \preccurlyeq_S \boldsymbol{S}_2.$$

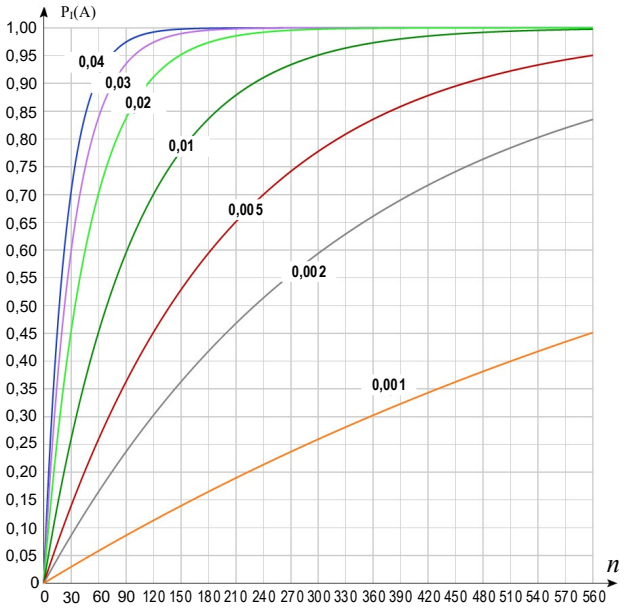


Рис. 2. Изменение характера зависимости $P_I(A)$ вероятности индицирования инцидента от числа наблюдаемых событий безопасности n при различных вероятностях p(0,001 → 0,04) возникновения этих событий

Fig. 2. The Changing of Dependence P_I (A) Probability of Incident Recognition as a Function of the Number of Observed Security Events n for Various Probabilities p (0,001 → 0,04) of These Events Occurring

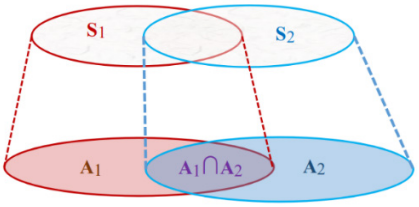


Рис. 3. Общий случай соотношения множества событий безопасности

Fig. 3. The General Case of the Relation of a Set of Security Events

Несмотря на невысокую точность оценок, описанный выше подход обладает примечательным свойством – все СЗИ сравнимы между собой. Однако допущение того (П-4), что значимость всех сигнатур одинакова для индицирования инцидента, является слишком сильным и далеко не всегда корректна, кроме того, введенное отношение позволяет сравнить СЗИ только для одного инцидента.

Отношение частичного порядка на множествах сигнатур событий безопасности для сравнения СЗИ с учетом значимости сигнатур

Более универсальным, более точным (снимающим ограничения предположения (П-4) о равной значимости событий безопасности), однако, трудоемким и, возможно, не всегда технически реализуемым (так как требуется набрать значительный объем статистики) методом является использование корреляционного анализа, применяемого в методах прикладной статистики [17]. Парные или частные коэффициенты корреляции можно рассматривать как коэффициенты значимости в процессе аддитивной свертки критериев, в отличие от

описанной выше свертки с одинаковыми коэффициентами значимости, равными 1. Использование этих коэффициентов возможно в качестве метрики для оценки связи между различными событиями безопасности, а также коэффициента множественной корреляции для оценки связи событий безопасности и инцидента. Т. е. в качестве оценки «степени связности» события безопасности и инцидента, в соответствии с их определением как зависимых событий, можно рассматривать корреляцию между ними. Тогда в общем случае связность события безопасности и инцидента определяется их парным коэффициентом корреляции.

Если происходит несколько событий безопасности (a_1, ..., a_n), индицирующих инцидент I_s, то показателем связности (тесноты) между этими событиями и инцидентом выступает коэффициент множественной корреляции $\boldsymbol{\rho}_{I\cdot a1,\ldots,an}$ [18]:

$$\boldsymbol{\rho}_{I\cdot a1,\ldots,an} = 1 - \frac{|R|}{R_{00}},$$

где R – корреляционная матрица $R = \{\boldsymbol{\rho}_{i,i}\}$; $|R|$ – определитель этой матрицы; R_{00} – алгебраическое дополнение для $\boldsymbol{\rho}_{0,0}$.

Сама корреляционная матрица определяется как:

$$R = \begin{pmatrix} \boldsymbol{\rho}_{0,0} & \cdots & \boldsymbol{\rho}_{0,an} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\rho}_{an,0} & \cdots & \boldsymbol{\rho} \end{pmatrix},$$

где $\boldsymbol{\rho}_{i,j} = \boldsymbol{\rho}_{ai,aj}$ – парные коэффициенты корреляции при $i, j = 1, \ldots, n$, а $\boldsymbol{\rho}_{0,0} = \boldsymbol{\rho}_{I,I}$ и все $\boldsymbol{\rho}_{i,i} = 1$.

Так как события безопасности независимы, то мультиколлериальность отсутствует (строки линейно-независимы): $\boldsymbol{\rho}_{i,j} = 0$ при $i \neq j$.

Множественный коэффициент корреляции можно использовать как метрику для оценки эффективности набора сигнатур $\boldsymbol{S}_k$:

$$\|\theta(\boldsymbol{S}_k)\| = \|\boldsymbol{A}_k\| = \boldsymbol{\rho}_{I\cdot\boldsymbol{A}_k} = \boldsymbol{\rho}_{I\cdot a1,\ldots,an}.$$

Часто, как показатель близости, также используется коэффициент детерминации:

$$D = (\boldsymbol{\rho}_{I\cdot\boldsymbol{A}_k})^2.$$

Необходимым условием корреляционного анализа является линейная регрессионная связь между объясняемым параметром и объясняющими факторами. Для рассматриваемой задачи объясняемым параметром является инцидент I, а объясняющими факторами – сигнатуры (агрегированные события безопасности). Как это предлагалось в [12], будем рассматривать все переменные этой линейной регрессии как дихотомические (бинарные), принимающие значения: 1 – если событие (инцидент) произошло; 0 – в противном случае.

При сравнении бинарных переменных, измеренных в дихотомической шкале, мерой корреляционной связи служит коэффициент ассоциации Пирсона φ. Величина коэффициента $\boldsymbol{\rho}_{i,j} = \boldsymbol{\rho}_{ai,aj} = \varphi$ лежит в интервале +1 и –1.

В общем виде формула эмпирического вычисления коэффициента корреляции дихотомических переменных $x = a_i$ и $y = a_j$ [19]:

$$\varphi = \frac{p_{xy} - p_x p_y}{\sqrt{p_x(1-p_x)p_y(1-p_y)}},$$

где p_x – оценка вероятности появления события $x = a_i$; p_y – оценка вероятности появления события $y = a_j$; p_{xy} – оценка вероятности появления событий x и y одновременно.

Более наглядно вычисление φ производится на основе таблицы 1 по результатам наблюдения, но с тем же результатом [19]:

$$\varphi = \frac{ad - bc}{\sqrt{(a+b)(b+d)(a+c)(c+d)}}.$$

ТАБЛИЦА 1. Таблица сопряжения

TABLE 1. Conjugacy Table

Событие a_i	Событие a_j		Σ
	Да	Нет	
Да	a	b	$a + b$
Нет	c	d	$c + d$
Σ	$a + c$	$b + d$	$n = a + b + c + d$

Известны также другие показатели связи, вычисляемые по этой таблице. Например, коэффициент ассоциации Юла [19]:

$$q = \frac{ad - bc}{ad + bc}.$$

В рассматриваемом случае асимптотического управления предполагается постепенное изменение СЗИ, в результате последовательного изменения множества контролируемых событий безопасности. Для оценки влияния отдельных факторов (в нашем случае событий безопасности) используются частные коэффициенты корреляции, когда значения всех остальных факторов фиксируются [19]. Метод последовательного включения (или исключения) событий безопасности позволяет выбрать из возможного набора событий именно те, которые повысят качество СЗИ. Например, это происходит, когда вводится новое событие безопасности a_i, имеющее наибольшее по абсолютной величине значение частного коэффициента корреляции с инцидентом при фиксированном влиянии ранее введенных событий безопасности:

$$\boldsymbol{\rho}_{I\cdot ai|a1,..,ai-1,ai+1,\ldots,an}.$$

В общем случае можно рассматривать включение (или исключения) группы событий безопасности, a_{i+1}, ..., a_{i+l} ($l < n$), тогда используется частный коэффициент корреляции l-го порядка.

При введении дополнительных событий безопасности коэффициент детерминации (множественной корреляции) должен возрастать (чем больше, тем лучше), а остаточная дисперсия уменьшаться. Следовательно, можем вести отношения предпочтения $\succcurlyeq_R$, сравнения сигнатур $\boldsymbol{S}_i$ на основе введенной метрики $\|\theta(\boldsymbol{S}_k)\| = \boldsymbol{\rho}_{I\cdot\boldsymbol{A}_k}$ и, соответственно, $\succcurlyeq_R$ для сравнения СЗИ:

$$\text{СЗИ}_1 \succcurlyeq_R \text{СЗИ}_2, \text{ если } \boldsymbol{S}_1 \succcurlyeq_R \boldsymbol{S}_2;$$

$$\text{СЗИ}_1 \preccurlyeq_R \text{СЗИ}_2, \text{ если } \boldsymbol{S}_1 \preccurlyeq_R \boldsymbol{S}_2;$$

$$\text{СЗИ}_1 =_R \text{СЗИ}_2, \text{ если } \boldsymbol{S}_1 =_R \boldsymbol{S}_2.$$

Вместо заключения

Рассмотренные варианты отношения порядка позволяют сравнивать между собой СЗИ и тем самым оценивать улучшение свойств СЗИ в процессе асимптотического управления. И если использование отношения $\succcurlyeq_R$ представляется более точным, то для $\succcurlyeq_S$ можно отметить его простоту и универсальность.

Использование метрики $||\theta(\boldsymbol{S}_r)||=||\boldsymbol{A}_r||$ (введенной для $\succcurlyeq_S$ или $\succcurlyeq_R$) в качестве универсальной ограничивает то, что описанные выше варианты отношения порядка вполне корректны только для случая одного инцидента.

В случае множества инцидентов можно рассматривать инцидент $\boldsymbol{I} = \{\boldsymbol{I}_1, \boldsymbol{I}_2, ..., \boldsymbol{I}_m\}$ как набор инцидентов одинаковой значимости, обнаруживаемых СЗИ, и сравнение СЗИ должно учитывать эффективность обнаружения каждого. Т. е. мы можем рассматривать это как задачу многокритериального выбора, где отдельным критерием выступает эффективность обнаружения каждого $\boldsymbol{I}_r$ (т. е. снимаются ограничения предположения П-1 и П-3). Для сравнения СЗИ по множествам сигнатур соответственно будет использоваться векторный критерий $(\boldsymbol{S}_1, \boldsymbol{S}_2, ..., \boldsymbol{S}_m)$.

Для векторных критериев возможны разные подходы, в частности можно использовать свертки (агрегирование) исходных данных [20]. Так, введенное отношение $\succcurlyeq^S$ или $\succcurlyeq^R$ можно распространить на общий случай для множества инцидентов индицируемых в СЗИ, если в качестве единого инцидента $\boldsymbol{I}$ рассматривать возникновение любого инцидента. В общем случае сравнение производится только как сравнение векторов одинаковой размерности $(\boldsymbol{S}_1^i, \boldsymbol{S}_2^i, ..., \boldsymbol{S}_m^i)$ и $(\boldsymbol{S}_1^j, \boldsymbol{S}_2^j, ..., \boldsymbol{S}_m^j)$, и $(\boldsymbol{S}_1^i, \boldsymbol{S}_2^i, ..., \boldsymbol{S}_m^i) \succcurlyeq (\boldsymbol{S}_1^j, \boldsymbol{S}_2^j, ..., \boldsymbol{S}_m^j)$ если $\boldsymbol{S}_r^i \succcurlyeq_S \boldsymbol{S}_r^j$ для любого r (или $\boldsymbol{S}_r^i \succcurlyeq_R \boldsymbol{S}_r^j$), в противном случае вектора несравнимы.

Однако следует учесть, что при асимптотическом управлении мы улучшаем качество СЗИ постепенно, последовательно внося изменения и, следовательно, можно предположить, что в большинстве случаев это будет связано с одним вполне конкретным инцидентом.

В заключение следует отметить, что, учитывая возможность динамики развития состава индицируемых инцидентов, следует одновременно предполагать неизменность защищаемого объекта и среды (так как их изменение может сокращать множество инцидентов). В частности, следует допускать, что переход от одной модели угроз к другой, основанный на изменении состава инцидентов, происходит при неизменном составе механизмов защиты.

Список используемых источников

1. Хоффман Л. Современные методы защиты информации. Пер. с англ. М.: Сов. радио, 1980. 262 с.
2. ГОСТ Р ИСО/МЭК 15408-1-2013 Информационная технология (ИТ). Методы и средства обеспечения безопасности. Критерии оценки безопасности информационных технологий. Часть 1. Введение и общая модель. М.: Стандартинформ, 2014.
3. Захарченко Р.И., Королев И.Д. Методика оценки устойчивости функционирования объектов критической информационной инфраструктуры, функционирующей в киберпространстве // Наукоемкие технологии в космических исследованиях Земли. 2018. Т. 10. № 2. С. 52–61. DOI:10.24411/2409-5419-2018-10041
4. Mikhalevich I.F., Trapeznikov V.A. Critical Infrastructure Security: Alignment of Views // Systems of Signals Generating and Processing in the Field of on Board Communications (Moscow, Russia, 20–21 March 2019). IEEE, 2019. DOI:10.1109/SOSG.2019.8706821
5. Ерохин С.Д., Петухов А.Н., Пилюгин П.Л. Принципы и задачи асимптотического управления безопасностью критических информационных инфраструктур // T-Comm: Телекоммуникации и транспорт. 2019. Т. 13. № 12. С. 29–35. DOI:10.24411/2072-8735-2018-10330
6. ISO/IEC 27005:2018(en) Information technology. Security techniques. Information security risk management // Online Browsing Platform. URL: https://www.iso.org/obp/ui/#iso:std:iso-iec:27005:ed-3:v1:en (дата обращения 07.09.2020)
7. Гавдан Г.П., Иваненко В.Г., Салкуцан А.А. Обеспечение безопасности значимых объектов критической информационной инфраструктуры // Безопасность информационных технологий. 2019. Т. 26. № 4. С. 69–82. DOI:10.26583/bit.2019.4.05
8. Ерохин С.Д., Петухов А.Н., Пилюгин П.Л. Событийно-ориентированная политика безопасности и формальная модель механизма защиты критических информационных инфраструктур // Труды учебных заведений связи. 2019. Т. 5. № 4. С. 99–105. DOI:10.31854/1813-324X-2019-5-4-99-105
9. Ерохин С.Д., Петухов А.Н., Пилюгин П.Л. Эффективность активного мониторинга событий сетевой безопасности // Электросвязь. 2020. № 2. С. 46–51. DOI:10.34832/ELSV.2020.3.2.007
10. Методические рекомендации по категорированию объектов критической информационной инфраструктуры, принадлежащих субъектам критической информационной инфраструктуры, функционирующим в сфере связи. АДЭ, 26.06.2019. URL: http://www.rans.ru/images/metrecKII.pdf (дата обращения 07.09.2020)

11. ГОСТ Р ИСО/МЭК ТО 18044-2007 Информационная технология (ИТ). Методы и средства обеспечения безопасности. Менеджмент инцидентов информационной безопасности. М.: Стандартинформ, 2009.
12. Масич Г.Ф. Системы обнаружения вторжений. Intrusion Detection System – IDS // ИМССУрОРАН. 2016. URL: https://www.icmm.ru/uchebnaya-deyatelnost/lektsii/514-ids (дата обращения 07.09.2020)
13. Ложковский А.Г., Каптур В.А., Вербанов О.В., Колчар В.М. Математическая модель пакетного трафика // Вестник Нац. техн. ун-та «ХПИ». 2011. № 9. С. 113–119. URL: http://repository.kpi.kharkov.ua/handle/KhPI-Press/10831 (дата обращения 07.09.2020)
14. Шелухин О.И. Сетевые аномалии. Обнаружение, локализация, прогнозирование. Москва: Горячая линия–Телеком, 2019. 448 с.
15. Новикова Е.С., Бекенева Я.А., Шоров А.В., Федотов Е.С. Обзор алгоритмов корреляции событий безопасности для обеспечения безопасности облачных вычислительных сред // Информационно управляющие системы. 2017. № 5(90). С. 95–104. DOI:10.15217/issn1684-8853.2017.5.95
16. Петухов А.Н., Пилюгин П.Л. Методы и инструменты моделирования безопасности критических информационных инфраструктур // Вторая всероссийская конференция «Современные технологии обработки сигналов» (СТОС-2019, Москва, Россия, 11–12 декабря 2019). Москва: Московское НТО радиотехники, электроники и связи им. А.С. Попова, 2019.
17. Фёрстер Э., Рёнц Б. Методы корреляционного и регрессионного анализа. Руководство для экономистов. Пер. с нем. М.: Финансы и статистика, 1983. 304 с.
18. Крамер Г. Математические методы статистики. Пер. с англ. М.: Мир, 1975. 648 с.
19. Ермолаев О.Ю. Математическая статистика для психологов. М.: Московский психолого-социальный институт, Флинта, 2003. 336 с.
20. Подиновский В.В. Идеи и методы теории важности критериев. М: Наука, 2019. 103 с.

* * *

About the Comparison of Information Security Systems for Asymptotic Information Security Management of Critical Information Infrastructures

S. Erokhin[1], **A. Petukhov**[1, 2*], **P. Pilyugin**[1, 3]

[1]Moscow Technical University of Communications and Informatics,
Moscow, 111024, Russian Federation
[2]National Research University of Electronic Technology (MIET),
Moscow, 124498, Russian Federation
[3]Lomonosov Moscow State University
Moscow, 119991, Russian Federation

Article info
DOI:10.31854/1813-324X-2020-6-3-66-74
Received 19th August 2020
Accepted 8th September 2020

For citation: Erokhin S., Petukhov A., Pilyugin P. About the Comparison of Information Security Systems for Asymptotic Information Security Management of Critical Information Infrastructures. *Proc. of Telecom. Universities.* 2020;6(3):66–74. (in Russ.) DOI:10.31854/1813-324X-2020-6-3-66-74

Abstract: *The article discusses the security management capabilities of critical information infrastructures. It discusses approaches to developing security policies that don't lean on assessing residual risks and identifying a fixed list of threats. We examine the possibility of building information security management systems based on monitoring of security events. A formal description of security events as well as relevant protection methods is proposed. The paper introduces an order relation for information security systems comparison and asymptotic CII security control implementation.*

Keywords: *information security, information security policy, information security events, information security events monitoring, asymptotic management, network events characteristics, critical information infrastructures.*

References

1. Khoffman L. *Modern Methods of Information Protection*. Moscow: Sov. radio Publ; 1980. 262 p. (in Russ.)
2. GOST R ISO/MEK 15408-1-2013 I*nformation technology. Security techniques. Evaluation criteria for IT security. Part 2. Security functional components.* Moscow: Standartinform Publ.; 2014. (in Russ.)
3. Zakharchenko R.I., Korolev I.D. Methods of Estimation of Stability of Functioning of Objects of Critical Information Infrastructure Operating in Cyberspace. *H&ES RESEARCH*. 2018;10(2):52–61. (in Russ.) DOI:10.24411/2409-5419-2018-10041
4. Mikhalevich I.F., Trapeznikov V.A. Critical Infrastructure Security: Alignment of Views. *Proceedings of the Conference on Systems of Signals Generating and Processing in the Field of on Board Communications, 20–21 March 2019, Moscow, Russia*. IEEE; 2019. DOI:10.1109/SOSG.2019.8706821
5. Erokhin S.D., Petukhov A.N., Pilyugin P.L. Principles and Tasks of Asymptotic Security Management of Critical Information Infrastructures. *T-Comm*. 2019;13(12):29–35. (in Russ.) DOI:10.24411/2072-8735-2018-10330
6. *Online Browsing Platform*. ISO/IEC 27005:2018(en) Information technology. Security techniques. Information security risk management. Available from: https://www.iso.org/obp/ui/#iso:std:iso-iec:27005:ed-3:v1:en [Accessed 7th September 2020]
7. Gavdan G.P., Ivanenko V.G., Salkutsan A.A. Security of Significant Objects of Critical Information Infrastructure. *IT Security*. 2019;26(4):69–82. (in Russ.) DOI:10.26583/bit.2019.4.05
8. Erokhin S., Petukhov A., Pilyugin P. Event-based Security Policy and Formal Model of Critical Information Infrastructures Protecting Mechanism. *Proc. of Telecom. Universities.* 2019;5(4):99–105. (in Russ.) DOI:10.31854/1813-324X-2019-5-4-99-105
9. Erokhin S.D., Petukhov A.N., Pilyugin P.L. Effectiveness of Active Monitoring of Network Security Events. *Elektrosviaz*. 2020;2:46–51. (in Russ.) DOI:10.34832/ELSV.2020.3.2.007
10. *Methodical Recommendations for Categorizing Critical Information Infrastructure Objects Belonging to Subjects of Critical Information Infrastructure Operating in the Field of Communications*. Documentary Telecommunication Association, 26th June 2019. (in Russ.) Available from: http://www.rans.ru/images/metrecKII.pdf [Accessed 7th September 2020]
11. GOST R ISO MEK/TO 18044-2007 *Information technology. Security techniques. Information security incident management*. Moscow: Standartinform Publ.; 2014. (in Russ.)
12. Masich G.F. Intrusion Detection System – IDS. *Institute of Continuous Media Mechanics of the Ural Branch of Russian Academy of Science*. 2016. (in Russ.) Available from: https://www.icmm.ru/uchebnaya-deyatelnost/lektsii/514-ids [Accessed 7th September 2020]
13. Lozhkovskii A.G., Kaptur V.A., Verbanov O.V., Kolchar V.M. Mathematical Model of Packet Traffic. *Vestnik Nat. tech. university "KhPI"*. 2011;9:113–119. (in Russ.) Available from: http://repository.kpi.kharkov.ua/handle/KhPI-Press/10831 [Accessed 7th September 2020]
14. Shelukhin O.I. *Network Anomalies. Detection, Localization, Forecasting*. Moscow: Goriachaia liniia–Telekom Publ.; 2019. 448 p. (in Russ.)
15. Novikova E.S., Bekeneva Ya.A., Shorov A.V., Fedotov E.S. A Survey of Security Event Correlation Techniques for Cloud Computing Environment Security. *Information and Control Systems*. 2017;5(90):95–104. (in Russ.) DOI:10.15217/issn1684-8853.2017.5.95
16. Petukhov A.N., Pilyugin P.L. Methods and Tools for Modeling the Security of Critical Information Infrastructures. *Proceedings of 2nd All-Russian Conference on Modern Signal Processing Technologies, 11–12 December 2019, Moscow, Russia*. Moscow: Moscow Scientific and Technical Society Radio Engineering, Electronics and Communications named after A.S. Popov Publ.; 2019. (in Russ.)
17. Förster E., Rönz B. *Methods of Correlation and Regression Analysis*. Moscow: Finansi i statistika Publ.; 1983. 304 p. (in Russ.)
18. Kramer G. *Mathematical Methods of Statistics*. Moscow: Mir Publ.; 1975. 648 p. (in Russ.)
19. Ermolaev O.Yu. *Mathematical Statistics for Psychologists*. Moscow: Moscow Psychological and Social Institute Publ., Flinta Publ.; 2003. 336 p. (in Russ.)
20. Podinovskiy V.V. *Ideas and Methods of the Theory of the Importance of Criteria*. Moscow: Nauka Publ.; 2019. 103 p. (in Russ.)

Сведения об авторах:

ЕРОХИН Сергей Дмитриевич — кандидат технических наук, доцент, ректор Московского технического университета связи и информатики, esd@mtuci.ru
https://orcid.org/0000-0001-8449-3612

ПЕТУХОВ Андрей Николаевич — кандидат технических наук, начальник отдела НИЧ Московского технического университета связи и информатики, anpetukhov@yandex.ru
https://orcid.org/0000-0002-1427-2440

ПИЛЮГИН Павел Львович — кандидат технических наук, старший научный сотрудник НИЧ Московского технического университета связи и информатики, ppl@mail.ru
https://orcid.org/0000-0003-0011-7180

УДК 004.453; 519.854 DOI:10.31854/1813-324X-2020-6-3-75-90

Построение комплексных расписаний обработки пакетов данных в конвейерной системе при задании ограничений на длительность интервалов времени ее функционирования

К.В. Кротов[1*]

[1]Севастопольский государственный университет,
Севастополь, 299053, Российская Федерация
[*]Адрес для переписки: krotov_k1@mail.ru

Информация о статье
Поступила в редакцию 18.02.2020
Принята к публикации 01.09.2020

Ссылка для цитирования: Кротов К.В. Построение комплексных расписаний обработки пакетов данных в конвейерной системе при задании ограничений на длительность интервалов времени ее функционирования // Труды учебных заведений связи. 2020. Т. 6. № 3. С. 75–90. DOI:10.31854/1813-324X-2020-6-3-75-90

Аннотация: *Рассматривается задача планирования обработки пакетов данных в конвейерной системе при ограничении на длительности интервалов времени ее функционирования. Решение задачи предполагает определение составов пакетов данных, составов групп пакетов данных, обрабатываемых в течение этих интервалов времени, расписаний обработки пакетов каждой из групп. Для оптимизации решений применен аппарат теории иерархических игр. Введены условия, позволяющие определять пакеты, обработка либо переналадка на обработку которых вызывает максимальные простои сегментов конвейера. Предложен метод построения эффективных составов групп, предполагающий исключение из них пакетов, определяемых в соответствии с этими условиями, и размещение в группах не вошедших в них пакетов.*

Ключевые слова: *иерархическая игра, интервалы функционирования конвейерной системы, группы пакетов данных, локальная оптимизация.*

Введение

Развитием иерархического подхода по определению комплексных расписаний обработки пакетов данных (ПД) [1] является решение задачи построения комплексных расписаний обработки пакетов при задании ограничения на интервалы времени функционирования конвейерной системы (КС). В этом случае обработка ПД выполняется в течение интервалов времени заданной длительности. Разделяемым ресурсом является ограниченное время функционирования КС, составы пакетов и расписания их обработки должны быть определены таким образом, чтобы этот ресурс был использован в максимальной степени.

С целью реализации обработки ПД в течение временных интервалов заданной длительности требуется сформировать их непересекающиеся множества, в каждое из которых входят ПД, обрабатываемые в течение одного из этих временных интервалов. В этом случае для планирования обработки ПД в КС требуется выполнить формирование следующих решений: по составам ПД, по составам непересекающихся множеств, каждое из которых содержит пакеты, обрабатываемые в течение одного из временных интервалов, а также решений по порядкам обработки ПД, входящих в каждое из указанных непересекающихся множеств пакетов, на сегментах КС.

Для решения комплексной задачи планирования обработки ПД в КС при задании ограничения на длительности интервалов времени ее функционирования выполнена декомпозиция общей цели системы построения комплексных расписаний на совокупность иерархически взаимосвязанных подфункции, каждая из которых реализуется на соответствующем уровне. На верхнем (первом) уровне реализуется формирование решений по составам ПД, на втором уровне формируются совокупности пакетов, обрабатываемые в течение временных интервалов заданной длительности, на третьем уровне реализуется построение расписаний обработки ПД в КС. С целью построения дальнейших рассуждений в рассмотрение введено понятие группы ПД, обрабатываемых в течение одного из временны́х интервалов.

1. Анализ существующих методов построения комплексных расписаний обработки ПД в КС

Развитие современных моделей и методов комплексного планирования обработки ПД (построения расписаний обработки ПД) представлено в многочисленных работах, в которых введено понятие задания и рассматриваются методы формирования пакетов заданий (ПЗ) и расписаний их выполнения.

В работах [2–6] рассмотрены методы планирования обработки партий материалов (ПМ) в системах типа *FlowShop* и *JobShop*, реализующих непрерывные технологические процессы получения продуктов производства. Количество продукции разного вида определяется потребностями рынка (объем выпуска является заданным). В соответствии с технологией производства определяются размеры обрабатываемых ПМ, порядок действий с ними на приборах системы. Для решения задачи планирования применен аппарат смешанного целочисленного программирования, что позволяет найти решение рассматриваемой задачи ограниченной размерности [2, 4].

В [7–10] рассмотрены методы построения расписаний выполнения фиксированных ПЗ на одном или нескольких параллельно действующих приборах. Для решения задачи построения расписания выполнения ПЗ на одном приборе в [7] введена формула, определяющая значения приоритета для каждого пакета на основе значений длительностей выполнения заданий в нем и длительностей переналадки приборов на выполнения заданий этого типа. Полученные значения введенного приоритета используются при упорядочивании ПЗ в очереди на выполнение.

Также в [7] предложен способ решения задачи определения составов ПЗ и расписаний их выполнения на нескольких параллельно действующих приборах, основывающийся на аппарате динамического программирования (ДП). Однако его применение ограничено размерностью задачи.

В [8] рассматривается задача построения расписаний выполнения ПЗ на одном обрабатывающем приборе, состоящем из параллельно действующих устройств. Пакеты подаются на прибор для реализации действий в порядке убывания потребности в результатах выполнения. Выполнение ПЗ на приборе упорядочивается в соответствии с убыванием этой потребности в результатах.

Решению задачи определения порядка выполнения фиксированных ПЗ на одном приборе, состоящем из параллельно действующих устройств, посвящены работы [9, 10]. В них рассматриваются два альтернативных подхода к построению расписания выполнения фиксированных ПЗ: на основе значений приоритета, которые вычисляются с учетом веса ПЗ и длительности выполнения заданий, и на основе построения полного графа выполнения ПЗ и определения на нем кратчайших путей между вершинами (в соответствии с найденными таким образом путями формируется порядок выполнения ПЗ).

В [11] рассматривается решение задачи оптимизации составов ПЗ, выполняемых на одном приборе, с учетом директивных сроков, формируемых для заданий (обрабатывающим прибором является средство доставки грузов, а ПЗ – совокупность грузов). Грузы упорядочиваются с точки зрения их директивных сроков, после чего из них формируются пакеты с учетом директивных сроков грузов и пропускной способности обрабатывающего прибора (грузоподъемности транспортного средства).

Работа [12] посвящена решению задач планирования выполнения ПЗ на параллельно действующих приборах. Размер формируемого пакета не может превышать количество параллельно действующих приборов. Особенностью задачи также является задание директивных сроков для выполняемых заданий. Критерием является количество ПЗ, которое должно быть минимизировано. Все задания с точки зрения длительностей выполнения являются разнотипными, поэтому объединение их в пакеты реализуются с точки зрения директивных сроков. Метод формирования решений предполагает, что из всех заданий, входящих в ПЗ, предшествующие некоторому рассматриваемому пакету, выбираются задания с максимальным временем завершения. Эти задания добавляются в рассматриваемый пакет. В результате задания из некоторого ПЗ могут быть распределены по другим пакетам, а количество ПЗ будет сокращено.

В работе [13] рассматривается решение задачи распределение заданий по пакетам и назначения сформированных ПЗ на параллельно действующие приборы с определением очередности их выполнения. Для решения задачи реализована ее декомпозиция на совокупность подзадач:

1) формирование составов ПЗ и их распределения по параллельно действующим приборам;

2) упорядочивание ПЗ в очередях на соответствующих приборах.

Решение первой подзадачи реализуется совместно с использованием эвристического правила, названного Apparent Tardiness Cost (*пер. с англ.* издержка видимой задержки, ИВЗ). Это правило предусматривает вычисление для заданий каждого типа значения параметра ИВЗ, учитывающего вес задания, длительность его выполнения, среднюю длительность выполнения заданий, не распределенных по приборам, момент времени поступления и т. д. Задания сортируются по не убыванию значения ИВЗ, после чего количество заданий, соответствующее размеру пакета, закрепляется за определенным незанятым прибором. Пакеты, назначенные для выполнения на приборы, упорядочиваются в очередях обработки с использованием генетиче-

ских алгоритмов (ГА). Анализ предложенного в [13] метода показал, что распределение заданий по пакетам выполняется на основе эвристического параметра ИВЗ и не оптимизируется.

В работе [14] также рассматривается применение эвристического подхода при определении составов ПЗ. Эвристическое правило формирования ПЗ предполагает, что в пакет включаются задания с близкими значениями длительностей. Для этого задания предварительно сортируются по времени выполнения, а затем распределяются по пакетам с учетом ограничений на вес и размер. Группирование заданий таким образом обеспечит минимизацию общего времени выполнения всех заданий. В работе [15] также развивается эвристический подход. Формулируемые в [15] правила положены в основу процедуры формирования решений по составам ПЗ и расписаниям их выполнения. Каждое эвристическое правило позволяет формировать одну из окрестностей текущего локально оптимального решения с заданной метрикой. В этих окрестностях выполняется определение нового комплексного (по составам ПЗ и расписаниям) решения.

Эвристическое правило определения составов ПЗ, выполняемых на параллельных машинах, используется и в методе, рассмотренном в [16]. Метод предполагает, что отношение предшествования между заданиями, поступающими в систему, определяющее порядок их выполнения, является заданным. Формирование ПЗ осуществляется по правилу: для каждого момента времени планирования в новый формируемый пакет включаются доступные не пакетированные задания, которые не связаны отношением предшествования с заданиями, не включенными в сформированные пакеты).

Рассмотрению способа формирования ПЗ с использованием функции стоимости, характеризующей эффективность использования прибора, состоящего из параллельных обрабатывающих устройств, посвящена работа [17]. Особенностью решаемой задачи является периодичность пакетной обработки и динамический характер планирования с учетом поступления заданий. Планирование осуществляется в ключевые моменты времени функционирования системы, которые соответствуют поступлению заданий в систему. Для каждого определенного указанным образом момента времени формируются составы ПЗ, которые характеризуются значениями стоимости использования системы. Стоимость использования системы для сформированного ПЗ вычисляется на основе значений параметров: количества заданий в формируемом пакете, пропускной способности системы, длительности выполнения заданий и текущего времени планирования. Пакет с максимальным значением стоимости передается для выполнения на прибор. Данный способ формирования ПЗ может быть применен только при незначительном количестве заданий в очереди на выполнение, иначе длительность формирования эффективных составов ПЗ будет значительной (т. к. необходим перебор большого количество решений).

Использованию метаэвристических алгоритмов при формировании решений по составам ПЗ, выполняемых на параллельных машинах, посвящены работы [18, 19]. Рассматривается задача построения расписания выполнения ПЗ на одном обрабатывающем приборе, состоящем из параллельно действующих устройств. Для определения составов ПЗ используются ГА и метод муравьиной колонии. Выбор наилучшего решения осуществляется с учетом требования минимизации простоев устройств прибора при выполнении ПЗ, которые должны быть сформированы из заданий, близких по длительностям обработки на устройствах. В силу стохастического характера ГА предложенный метод не может гарантировать получение эффективных решений при различных значениях входных параметров. Метод муравьиной колонии также не гарантирует получение эффективных решений при различных значениях входных параметров.

В [20] решается задача пакетирования выполнения заданий на параллельных машинах с использованием частично целочисленного линейного программирования (ЧЦЛП). В качестве решения определяются значения бинарных переменных, характеризующих включение заданий рассматриваемых типов в формируемые пакеты, а также местоположение заданий в обрабатывающем приборе. Множество решений сформировано введенной в рассмотрение системой ограничений (их общее количество равно 23). В работе [20] отмечается, что решение задачи планирования с использованием приведенной модели и ЧЦЛП возможно только в случае малых масштабов задачи (ограниченное количество заданий и выполняющих их машин и т. д.). Решение задачи при больших значениях этих параметров за ограниченное время является затруднительным.

Метод, использующий ЧЦЛП, рассматривается в работе [21]. Поиск эффективного решения задачи предполагает определение комплексного расписания выполнения ПЗ, включающего решения по составам пакетов и расписаниям. Решение задачи планирования предполагает определение значений непрерывных и бинарных переменных: моментов времени начала выполнения ПЗ; моментов времени завершения выполнения заданий в ПЗ; переменных, определяющих включение заданий в пакеты; переменных, определяющих отношение предшествования для пакетов в очереди на выполнение и т. д. В силу того, что для решения задачи используется аппарат ЧЦЛП, ее размерность является ограниченной (2–3 заказа, 2–3 задания в заказе, 2–3 параллельных устройства в обрабатывающем приборе).

В [22] рассматривается модель формирования составов ПЗ с точки зрения экономических показателей при учете спроса на продукцию, но без учета технологических характеристик процесса выполнения заданий. Решение задачи планирования предполагает формирование решений двух подзадач: среднесрочного планирования (распределение выполнения заданий (ПЗ) по предприятиям, выпускающим продукцию) и краткосрочного планирования (составление расписаний выполнения заданий на технологическом оборудовании предприятий). Распределение заданий по пакетам реализуется отдельно от распределения выполнения ПЗ на обрабатывающих приборах.

Анализ существующих методов комплексного планирования выполнения ПЗ позволил определить следующие их недостатки:

– в силу значительных вычислительных затрат использование методов ЧЦЛП возможно при малой размерности решаемой задачи;

– использование методов определения составов ПЗ на основе значений функций приоритета позволяет получить решения, лучшие, чем решения без оптимизации, но не приближающиеся к оптимальным;

– определения составов ПЗ на основе значений директивных сроков окончания заданий возможно только для ограниченного круга задач;

– применение методов определения составов ПЗ и расписаний их выполнения на основе эвристических правил, также, как и метаэвристических алгоритмов, не гарантирует получения эффективных решений.

Анализ методов комплексного планирования показал, что ни один из них не решает задачу определения составов ПЗ при учете ограничения на интервалы времени функционирования системы; также отсутствуют методы определения групп ПЗ, выполняемых в течение этих интервалов. В силу сказанного разработка методов комплексного планирования выполнения ПЗ с учетом ограничений, включающих методы определения эффективных составов ПЗ, составов групп ПЗ, расписаний их выполнения, является актуальной.

Излагаемые в данной статье результаты исследований базируются на предыдущих работах автора [1, 23, 24], в которых рассмотрены:

– метод формирования составов ПД и расписаний их обработки в КС [1], который позволяет получать комплексные эффективные (локально оптимальные) решения без наличия ограничений на интервалы времени функционирования КС;

– математическая модель иерархической игры, представляющая собой систему критериев принятия решений по составам ПД, составам групп ПД и расписаниям их обработки, позволяющая задавать порядок оптимизации решений, а также способ оценивания решений на вышестоящих уровнях иерархии на основе решений с нижестоящих уровней [23];

– метод формирования начальных решений по составам групп ПД, обрабатываемых в течение заданных интервалов времени функционирования КС, а также способ оптимизации решений по составам групп ПД с использованием метаэвристических алгоритмов (в частности, ГА) [24].

2. Математическая модель иерархической игры оптимизации комплексных решений по расписаниям обработки ПД в КС при задании ограничений и метод формирования начальных решений по составам групп ПД

Для решения задачи планирования обработки ПД в КС при задании ограничения на интервалы времени ее функционирования выполнена декомпозиция обобщенной функции системы построения комплексных расписаний на совокупность иерархически упорядоченных подфункций. Каждая из них реализуется с использованием соответствующей подсистемы принятия решений. Функционирование этих подсистем предполагает определение следующих решений:

– решений по составам ПД, обрабатываемых в КС (первый уровень);

– решений по составам групп ПД, обрабатываемых в течение временных интервалов заданной длительности (второй уровень);

– решений по порядкам обработки ПД из групп на сегментах КС – расписаниям обработки пакетов (третий (нижний) уровень).

С целью комплексной оптимизации решений на различных уровнях иерархии системы построения комплексных расписаний применен аппарат теории иерархических игр.

Для обоснования методов оптимизации комплексных расписаний обработки ПД в течение заданных временны́х интервалов функционирования системы введены следующие обозначения [1]: m_i – количество ПД i-го типа ($i = \overline{1,n}$), формируемых на первом уровне принятия решений, элементы m_i образуют вектор M; A – матрица, элементом a_{ih} которой является количество данных i-го типа в h-ом ПД ($h \le m_i$). Формируемое решение по составам ПД имеет вид: $[M, A]$. Решение задачи оптимизации составов групп ПД, обозначенных как N^z ($z = \overline{1,Z}$), предполагает задание интервалов времени t^z ($z = \overline{1,Z}$) функционирования системы при обработке данных.

В соответствии с обозначениями [23] решение, формируемое на втором уровне иерархии, представляется множеством $\{N^z | z = \overline{1,Z}\}$ наборов параметров вида: $N^z = \{[i, m_i^z, (A)_i^z]_k, | k = \overline{1,k_z}\}^z$, где m_i^z – количество ПД i-го типа, включенных в группу N^z; $(A)_i^z$ – вектор, каждый j-ый элемент которого соответствует составу одного из m_i^z ПД i-го типа (вектор $(A)_i^z$ – вектор составов ПД i-го типа, вклю-

ченных в группу N^z); k_z – количество типов данных, пакеты которых включены в группу N^z.

С целью учета ПД, не вошедших в состав групп N^z ($z = \overline{1,Z}$), введено множество Q – множество наборов параметров:

$$Q = \{[i, m_i^q, (A)_i^q]_k, |k = \overline{1,k_q}\},$$

где m_i^q – количество ПД i-го типа, не включенных в группы N^z; вектор $(A)_i^q$ – составы не обработанных ПД; k_q – количество типов данных, пакеты которых не включены в состав групп N^z.

Формируемое решение по составам групп ПД имеет следующий вид:

$$\{N^z|(z = \overline{1,Z})\} \text{ и } Q = \{[i, m_i^q, (A)_i^q]_k, |k = \overline{1,k_q}\},$$

где $N^z = \{[i, m_i^z, (A)_i^z]_k, |k = \overline{1,k_z}\}^z$).

Порядки обработки ПД группы N^z на сегментах конвейера представляются в виде последовательностей π^l реализации действий с ними на этих l-ых сегментах ($l = \overline{1,L}$). Расписание обработки ПД группы N^z формируется в предположении, что порядок реализации действий с ними в π^l является одинаковым на всех L сегментах конвейера.

С целью задания последовательностей π^l расписания π^z введена матрица порядка обработки ПД P^z [1] (размерность матрицы $k_z \times n_p^z$, где k_z – количество типов данных в ПД в группе N^z; n_p^z – число ПД в последовательностях π^l для группы N^z). В рассмотрение введена матрица R^z – матрица количества данных i-ых типов в ПД, занимающих в последовательностях π^l j-ые позиции [1], а также обозначения: $(t_i^l)^z$ – вектор длительностей обработки данных i-ых типов l-ом сегменте конвейера ($l = \overline{1,L}$), пакеты которых включены в группу N^z ($z = \overline{1,Z}$); $(t_{jq}^{0l})^z$ – матрица моментов времени начала обработки q-ых данных в пакетах, занимающих в π^l j-ые позиции ($z = \overline{1,Z}$, $l = \overline{1,L}$).

Расписание обработки ПД имеет вид:

$$\{[P^z, R^z, (t_{jq}^{0l})^z|(l = \overline{1,L})] \ |z = \overline{1,Z})\}.$$

В работе [23] выполнено обоснование модели иерархической игры, представляющей собой систему критериев на уровнях иерархии принятия решений.

Полученная в [23] модель иерархической игры имеет следующий вид:

– первый уровень, на котором определяются составы ПД: $\min f_1$, где:

$$f_1 = \sum_{i=1}^{n} n^i - \sum_{z=1}^{Z}\sum_{k=1}^{k_z}\sum_{h=1}^{[m_i^z]_k} [(a_h)_i^z]_k ;$$

– второй уровень, на котором определяются составы групп ПД: $\min f_2$, где:

$$f_2 = \sum_{z=1}^{Z}\left[\sum_{l=2}^{L}(t_{11}^{0l})^z + \sum_{l=1}^{L}\sum_{j=2}^{n_p^z}\left[(t_{j1}^{0l})^z - \left[(t_{j-1,n_{j-1}}^{0l})^z + \sum_{h=1}^{k_z}(t_h^l)^z \cdot p_{h,j-1}^z\right]\right] + \right.$$

$$\left. + \sum_{l=2}^{L}\sum_{j=1}^{n_p^z}\sum_{q=2}^{n_j^z}\left[(t_{jq}^{0l})^z - \left[(t_{j,q-1}^{0l})^z + \sum_{h=1}^{k_z}(t_h^l)^z \cdot p_{hj}^z\right]\right] + \left[\sum_{l=1}^{L}(t^z - [(t_{n_p^z,n_{n_p^z}}^{0l})^z + \sum_{h=1}^{k_z}(t_h^l)^z \cdot p_{h,n_p^z}^z]\right]\right] ;$$

– третий уровень, на котором определяются расписания обработки ПД из групп N^z на приборах КС: $\min f_3^z$, где:

$$f_3^z = \sum_{l=2}^{L}(t_{11}^{0l})^z + \sum_{l=1}^{L}\sum_{j=2}^{n_p^z}\left[(t_{j1}^{0l})^z - \left[(t_{j-1,n_{j-1}}^{0l})^z + \sum_{h=1}^{k_z}(t_h^l)^z \cdot p_{h,j-1}^z\right]\right] +$$

$$+ \sum_{l=2}^{L}\sum_{j=1}^{n_p^z}\sum_{q=2}^{n_j^z}\left[(t_{jq}^{0l})^z - \left[(t_{j,q-1}^{0l})^z + \sum_{h=1}^{k_z}(t_h^l)^z \cdot p_{hj}^z\right]\right] \text{ при } z = \overline{1,Z};$$

– ограничение для длительности реализации обработки ПД группы N^z, интерпретируемое для решений на третьем уровне:

$$\max_l\left[(t_{n_p^z,n_{n_p^z}}^{0l})^z + \sum_{h=1}^{k_z}(t_h^l)^z \cdot p_{h,n_p^z}^z\right] \le t^z, \text{ при } z = \overline{1,Z}. \quad (1)$$

Критерий на первом уровне определяет количество данных, не распределенных по группам N^z $(z = \overline{1,Z})$ (количество данных, не обработанных в течение заданных интервалов времени t^z $(z = \overline{1,Z})$). Критерий на втором уровне характеризует суммарные простои приборов КС при обработке ПД, входящих в группы N^z. Критерий на третьем уровне учитывает простои приборов КС для текущего количества пакетов, размещенных в последовательностях π^l при построении расписания [1].

На первом уровне иерархии системы планирования с целью оптимизации решений по составам ПД использован метод, рассмотренный в [1]. В основу способа построения расписаний обработки ПД положен метод, также предложенный в [1], но адаптация которого к рассматриваемой задаче предполагает рассмотрение только ПД, входящих в группы N^z $(z = \overline{1,Z})$. Метод определения эффективных решений по составам групп N^z $(z = \overline{1,Z})$ предполагает выполнение двух этапов:

1) формирование начального состава групп ПД;

2) переход к локально оптимальному решению по составам групп $\{N^{z*} | z = \overline{1,Z}\}$.

Формирование начального решения предполагает распределение ПД, полученных в решении $[M, A]$ с первого уровня, по множествам наборов параметров $N^z = \{[i, m_i^z, (A)_i^z]_k, | k = \overline{1,k_z}\}^z$ и множеству Q. Метод формирования начального решения по составам групп ПД, обрабатываемых в течение заданных временны́х интервалов функционирования КС, изложен в [24]. Формирование составов групп N^{z*} $(z = \overline{1,Z})$ на основе начального решения $\{N^z | z = \overline{1,Z}\}$ предполагает определение такого их состава, при котором суммарные простои сегментов конвейера при обработке ПД будут минимальными.

3. Метод определения составов групп ПД, обрабатываемых в течение заданных интервалов времени функционирования КС

Поиск локально оптимального решения по составам групп N^z $(z = \overline{1,Z})$ связан с определением в каждой группе ПД, обработка которого вызывает значительные простои сегментов, а также с исключением определенных таким образом ПД из соответствующих групп N^z $(z = \overline{1,Z})$ и размещением в этих группах ПД других типов, не вошедших в N^z (с учетом длительности интервалов t^z $(z = \overline{1,Z})$). Исключение из каждой группы N^z $(z = \overline{1,Z})$ ПД i-го типа, обработка которого обуславливает значительные простои сегментов, замещение его ПД другого i'-го типа $(i' \neq i)$, не вошедшего в состав групп N^z в решении $\{N^{z*} | z = \overline{1,Z}\}$, позволяет реализовать построение нового решения $\{N^z | z = \overline{1,Z}\}$. Тогда для $\{N^{z*} | z = \overline{1,Z}\}$ будет сформирована окрестность решений, в которой определяется новое эффективное решение $\{N^z | z = \overline{1,Z}\}$

Определение общих простоев сегментов конвейера при обработке ПД реализуется при зафиксированном (в соответствии с π^{z*}) порядке этих пакетов в $\pi^l (l = \overline{1,L})$. Поэтому идентифицируется (для исключения из группы N^z) не ПД определенного i-го типа, а ПД, занимающий некоторую j-ую позицию в последовательностях π^l $(l = \overline{1,L})$. Для идентификации ПД, исключаемого из N^z, необходимо определить суммарное время неэффективного использования ресурса всех сегментов конвейера при обработке этого пакета. Для построения рассуждений используем индекс j как номер позиции ПД в последовательностях π^l $(l = \overline{1,L})$ при условии, что $j \neq 1 (j \in \{2, 3, \ldots, n_p^z\})$. Интервал времени простоя l-го сегмента в ожидании начала обработки ПД, занимающего в $\pi^l j$-ую позицию, будет определен выражением вида [1]:

$$(t_{j1}^{0l})^z - \left[(t_{j-1,n_{j-1}^z}^{0l})^z + \sum_{h=1}^{k_z} (t_h^l)^z \cdot p_{h,j-1}^z\right]. \qquad (2)$$

Данный ПД располагается в π^l $(l = \overline{1,L})$ в j-ой позиции [1], тогда с использованием выражения (2) выполняется расчет интервалов времени простоев каждого сегмента в ожидании начала обработки этого пакета. Суммарное время простоев сегментов конвейера в ожидании обработки ПД, занимающего j-ую позицию в последовательностях π^l, будет вычислено в соответствии с выражением, где $j = \overline{2,n_p^z}$:

$$\sum_{l=1}^{L} [(t_{j1}^{0l})^z - [(t_{j-1,n_{j-1}^z}^{0l})^z + \sum_{h=1}^{k_z} (t_h^l)^z \cdot p_{h,j-1}^z]].$$

Обозначим номер позиции ПД в π^l (индекс ПД), исключаемого из группы N^z, через j_1. Тогда индекс j_1 позиции ПД в π^l, обработка которого в составе группы N^z вызывает максимальные простои сегментов конвейера в ожидании выполнения операций с ним, определяется выражением, где $j = \overline{2,n_p^z}$:

$$\max_j \left\{\sum_{l=1}^{L} [(t_{j1}^{0l})^z - [(t_{j-1,n_{j-1}^z}^{0l})^z + \sum_{h=1}^{k_z} (t_h^l)^z \cdot p_{h,j-1}^z]]\right\}. \qquad (3)$$

Интервал времени простоя l-го сегмента перед началом реализации действий с первым в π^l ПД определяется значением $(t_{11}^{0l})^z$. Тогда суммарное время простоя всех L сегментов перед началом реализации действий с первым в $\pi^l (l = \overline{1,L})$ ПД – выражением $\sum_{l=1}^{L} (t_{11}^{0l})^z$. На основе этого выражения и выражения (3) индекс j_1 (номер j-ой позиции ПД в π^l при $j = \overline{1,n_p^z}$), – определяется выражением:

$$j_1 = \begin{cases} 1, \text{если} \quad \sum_{l=1}^{L}(t_{11}^{0l})^z > \max_j\{\sum_{l=1}^{L}[(t_{j1}^{0l})^z - [(t_{j-1,n_{j-1}^z}^{0l})^z + \sum_{h=1}^{n_p^z}(t_h^l)^z \cdot p_{h,j-1}^z]]\}, \\ j', \text{если} \quad \sum_{l=1}^{L}(t_{11}^{0l})^z \le \max_j\{\sum_{l=1}^{L}[(t_{j1}^{0l})^z - [(t_{j-1,n_{j-1}^z}^{0l})^z + \sum_{h=1}^{n_p^z}(t_h^l)^z \cdot p_{h,j-1}^z]]\}. \end{cases} \quad (4)$$

Время простоя l-го сегмента в ходе обработки ПД, занимающего j-ую позицию в последовательности π^l (простои сегмента в ожидании готовности к обработке данных, входящих в пакет), определяется выражения вида [1]:

$$\sum_{q=1}^{n_j^z-1}[(t_{j,q+1}^{0l})^z - [(t_{jq}^{0l})^z + \sum_{h=1}^{k_z}(t_h^l)^z \cdot p_{h,j}^z]], \quad (5)$$

где n_j^z – количество данных в пакете, занимающем j-ую позицию в последовательностях π^l ($l = \overline{1,L}$), определяемое следующим образом: $n_j^z = \sum_{h=1}^{k_z} r_{hj}^z$.

Значение суммарных простоев сегментов конвейера в ожидании готовности к обработке данных в пакете, занимающем j-ую позицию в последовательностях π^l ($l = \overline{1,L}$) обозначим через t_j^z. В соответствии с выражением (5) значение t_j^z для каждого j-го ПД в π^l (j – номер позиции ПД в π^l) определяется выражением вида:

$$t_j^z = \sum_{l=1}^{L}\sum_{q=1}^{n_j^z-1}[(t_{j,q+1}^{0l})^z - [(t_{jq}^{0l})^z + \sum_{h=1}^{k_z}(t_h^l)^z \cdot p_{h,j}^z]]. \quad (6)$$

Через j_2 обозначим номер позиции ПД в π^l, обработка которого обуславливает максимальные простои сегментов конвейера в ожидании готовности данных в этом пакете. Значение индекса j_2 определяет позицию ПД (в последовательностях π^l), который должен быть исключен из группы N^z. С учетом выражения (6) значение индекса j_2 является таким, что для него выполняется условие вида (где $j = \overline{1, n_p^z}$):

$$t_{j_2}^z = \max_j[\sum_{l=1}^{L}\sum_{q=1}^{n_j^z-1}[(t_{j,q+1}^{0l})^z - [(t_{jq}^{0l})^z + \sum_{h=1}^{k_z}(t_h^l)^z \cdot p_{h,j}^z]], \quad (7)$$

Выражения (4) и (7) позволяют идентифицировать (по значению позиции j_1 или j_2 ПД) тип данных и их количество в пакетах в группе N^z, обработка которых обуславливает максимальные простои сегментов конвейера. Идентифицированные таким образом ПД могут быть удалены из группы N^z для замены пакетами из множества Q с целью уменьшения общих простоев сегментов конвейера, обрабатывающих пакеты группы N^z.

С целью обоснования метода построения эффективных составов групп ПД $\{N^{z*}|z = \overline{1,Z}\}$ в рассмотрение введены следующие обозначения:

p – индекс условия, в соответствии с которым реализуется определение ПД, исключаемого из группы $N^{z\prime}$: $p = 1$ соответствует условию (4); $p = 2$ соответствует условию (7));

D^z – вектора идентификаторов типов данных, пакеты которых размещены в группах $N^z (z = \overline{1,Z})$;

$j_p^{z\prime} (p = \overline{1,2})$ – номер позиции ПД в π^l расписания $\pi^{z\prime}$, обработка которого вызывает максимальные суммарные простои сегментов конвейера, определенные в соответствии с условиями (4) и (7): $p = 1$; $p = 2$;

$i_p^{z\prime} (p = \overline{1,2})$ – индекс типа данных, пакет которых извлекается из группы $N^{z\prime}$ (ПД $i_p^{z\prime}$-го типа, занимающий в π^l ($l = \overline{1,L}$) $j_p^{z\prime}$-ую позицию ($p = \overline{1,2}$), определяемую в соответствии с условиями (4) либо (7) ($p = \overline{1,2}$));

$r_p^{z\prime} (p = \overline{1,2})$ – количество данных в пакете, который извлекается из группы $N^{z\prime}$ (ПД $i_p^{z\prime}$-го типа, занимающий в π^l ($l = \overline{1,L}$) $j_p^{z\prime}$-ую позицию ($p = \overline{1,2}$));

$w_p (p = \overline{1,2})$ – счетчик количества ПД, входивших в множество Q, которые добавлены в группу $N^{z\prime}$ вместо исключаемого из нее ПД при формировании новых решений;

$i^{q\prime}$ – индекс типа данных в пакете, входящем в множество Q, который добавляется в группу $N^{z\prime}$ взамен исключаемого из этой группы ПД $i_p^{z\prime}$-го типа;

$r^{q\prime}$– количество данных в пакете $i^{q\prime}$-го типа, который добавляется в группу $N^{z\prime}$ взамен исключаемого из нее ПД $i_p^{z\prime}$-го типа;

$(i_w^q)_p$ $(w \in \{1,2,..,w_p\}_p, p = \overline{1,2})$ – вектора типов данных в пакетах, добавление которых в группу $N^{z\prime}$ взамен ПД $i_p^{z\prime}$-го типа $(p = \overline{1,2})$ обеспечивает минимальные отрицательные значения левых дискретных градиентов $(\nabla^- f_2)_p$ критерия f_2 для группы $N^{z\prime}$;

$(r_w^q)_p$ $(w \in \{1,2,..,w_p\}_p, p = \overline{1,2})$ – вектора количества данных в пакетах, добавление которых в группу $N^{z\prime}$ взамен ПД $i_p^{z\prime}$-го типа $(p = \overline{1,2})$ обеспечивает минимальные отрицательные значения левых градиентов $(\nabla^- f_2)$ критерия f_2;

$(\nabla^- f_2)_p$ $(p = \overline{1,2})$ – максимальные по модулю значения левых дискретных градиентов $(\nabla^- f_2)$ критерия f_2, получаемые для решений, в которых реализуется удаление из $N^{z\prime}$ ПД $i_p^{z\prime}$-ых типов в количестве $r_p^{z\prime}$ элементов $(p = \overline{1,2})$ и добавление в $N^{z\prime}$ данных $i^{q\prime}$-ых типов $(i_w^q = i^{q\prime})$ в количестве $r^{q\prime}$ $(r_w^q = r^{q\prime})$ элементов;

$(\nabla^- f_2)_G$ – «глобальное» значение левого дискретного градиента $(\nabla^- f_2)$ критерия f_2, которое соответ-

ствует наилучшему варианту решения в окрестности текущего локально оптимального решения $\{N^{z*} | \ z = \overline{1,Z}\}$;

$z1$ – сохраняемый номер группы N^z, обмен пакетами между которой и множеством Q обеспечивает значение $(\nabla^- f_2)_G$;

i^z – идентификатор типа данных в пакете, исключение которого из группы N^z ($z = z1$) обеспечивает значение $(\nabla^- f_2)_G$;

r^z – количество данных в пакете, исключение которого из группы N^z ($z = z1$) обеспечивает значение $(\nabla^- f_2)_G$;

$(i_w^q)_G$ $(w \in \{1, 2, .., w_p\})$ – вектор типов данных в пакетах, добавление которых в одну из групп N^z $(z = \overline{1,Z})$ взамен ПД i^z-го типа гарантирует значение $(\nabla^- f_2)_G$;

$(r_w^q)_G$ $(w \in \{1, 2, .., w_p\})$ – вектор количества данных в пакетах, добавление которых в одну из групп N^z $(z = \overline{1,Z})$ при исключении из нее ПД i^z-го типа гарантирует значение $(^- \nabla f_2)_G$;

$N_T^{z'}$ – промежуточное решение по составу группы $N^{z'}$, формируемое путем извлечения из нее ПД $i_p^{z'}$-го типа $(p = \overline{1,2})$;

$N_p^{z'}$ $(p = \overline{1,2})$ – решения, сформированные на основе промежуточного решения $N_T^{z'}$ по составу группы $N^{z'}$ путем добавления в эту группу ПД, входящих в множество Q;

H_{i^q} – множество номеров ПД i^q-го типа, которые могут быть добавлены в группу $N^{z'}$ вместо исключаемого из нее пакета (перед реализацией алгоритма определено k_q множеств H_{i^q}: $|H_{i^q}| = m_i^q$);

$H_{i^q}^T$– модифицируемое в ходе реализации алгоритма множество номеров ПД i^q-го типа, которые могут добавляться в группу $N^{z'}$ из множества Q.

Алгоритм определения локально оптимального решения по составам групп ПД $\{N^z | \ z = \overline{1,Z}\}$ оперирует со значениями индекса p, которым поставлены в соответствие условия (4) и (7). Он предполагает, что на основе текущих составов каждой группы N^z формируются два решения, каждое из которых предусматривает исключение из рассматриваемой группы N^z тех ПД, обработка которых обуславливает максимальные простои сегментов конвейера, определяемые условиями (4) и (7). В каждом из полученных решений реализуется дополнение группы N^z ПД различных типов, входящими в множество Q. Определение текущего локально оптимального решения по составу группы N^z предполагает, что в нее добавляется в первую очередь тот пакет, размещение которого обеспечивает максимальное улучшение значения критерия. Если для рассматриваемой группы N^z выполняется условие (1), то в нее добавляются ПД из множества Q. Дополнение группы N^z новыми ПД из множества Q продолжается до нарушения условия (1). Идентификаторами типов данных в пакетах, размещаемых в N^z, инициализируются элементы векторов $(i_w^q)_p$ $(p = \overline{1,2})$, которые сопоставлены с разными решениями, соответствующими ПД, исключаемыми из группы N^z при учете условий (4) и (7). Количество элементов в каждом из векторов равно w_p $(p = \overline{1,2})$. Количество данных в пакетах используется при инициализации элементов векторов $(r_w^q)_p$ $(p = \overline{1,2})$. Эти действия выполняются для всех групп $N^z (z = \overline{1,Z})$.

Последовательность шагов алгоритма оптимизации составов групп $N^z (z = \overline{1,Z})$ имеет следующий вид:

1) задание номера z' текущей рассматриваемой группы $N^{z'}$следующим образом: $z' = 1$ (группа $N^{z'}$, с которой будет выполняться обмен ПД из Q);

2) вычисление для решения $\{N^{z*} | \ z = \overline{1,Z}\}$ значения критерия $f_2(s)$ (для этого выполняется передача решения $\{N^{z*} | \ z = \overline{1,Z}\}$ на третий уровень иерархии системы, формирование расписаний $\pi^{z*}(z = \overline{1,Z})$, передача расписаний на второй уровень иерархии, вычисление значения $f_2(s)$ критерия f_2);

3) задание значения индекса p условия, в соответствии с которым в последовательностях π^l $(l = \overline{1,L})$ выделяются ПД с номерами позиций $j_p^{z'}$, равным 1 (p = 1); задание значений счетчиков w_p $(p = \overline{1,2})$ ПД, добавляемых в группу $N^{z'}$ из множества Q, равным 1 $(w_p = 1)$; задание значения $(\nabla^- f_2)_G = 0$;

4) выполняется на основе множества Q вида $Q = \{[i^q, m_i^q, (A)_i^q]_k | k = \overline{1,k_q}\}$, инициализация множеств H_{i^q} номеров ПД i^q-ых типов, которые могут быть размещены в группе $N^{z'}$; для этого реализуются следующие шаги:

– инициализация номера k рассматриваемого набора параметров в множестве Q: k = 1;

– инициализация значений элементов h_{i^q} вектора H_{i^q}: $H_{i^q} = \{1, 2, .., m_{i^q}\}$ для k-го набора;

– $k = k + 1$, проверка условия $k \le k_q$: если условие истинно, то переход на предыдущий шаг (инициализация значений элементов h_{i^q} вектора H_{i^q}); если условие ложно, то переход на шаг 5;

5) в соответствии с условием, идентифицируемым значением индекса p, определяется позиция $j_p^{z'}$ ПД в последовательностях π^l $(l = \overline{1,L})$, обработка которого вызывает максимальные простои сегментов (для этого задействовано решение с третьего уровня$\{[P^z, (t_{jq}^{0l})^z | (l = \overline{1,L})] \ | z = \overline{1,Z})\}$); в соответствии со значением $j_p^{z'}$ по матрице $P^{z'}$ идентифицируется номер строки i, для которой $p_{i,j_p^{z'}}^{z'} = 1$ при $(i = \overline{1,k_{z'}})$ (номер строки i соответствует номеру элемента вектора $D^{z'}$ типов данных, пакеты которых включены в состав группы $N^{z'}$); по значению i определение типа данных $i_p^{z'}$, пакет которых исключается из группы $N^{z'}$: $i_p^{z'} = d_i^{z'}$; в соответствии с номером строки i матрицы $P^{z'}$ идентификация по мат-

рице $R^{z'}$ количества данных $r^{z'}_{i,j^{z'}_p}$ в рассматриваемом ПД и инициализация параметра $r^{z'}_p$: $r^{z'}_p = r^{z'}_{i,j^{z'}_p}$;

6) в группе $N^{z'}$ по типу данных $i^{z'}_p$ идентифицируется набор $N_1 = [i^{z'}_p, m^{z'}_{i^{z'}_p}, (A)^{z'}_{i^{z'}_p}] \in N^{z'}$, с которым будут выполняться преобразования; инициализация промежуточного решения $N^{z'}_T$, на основании которого будет формироваться новое решение по составу группы $N^{z'}$: $N^{z'}_T = N^{z'}$; в решении $N^{z'}_T$ по составу ПД в группе $N^{z'}$ исключается ПД $i^{z'}_p$-го типа следующим образом:

– в наборе $N_1 \in N^{z'}$ для $i^{z'}_p$-го типа данных и их количества $r^{z'}_p$ определение в векторе $(A)^{z'}_{i^{z'}_p}$ индекса h' пакета такого, что $(a_{h'})^{z'}_{i^{z'}_p} = r^{z'}_p$;

– переупорядочивание элементов $(A)^{z'}_{i^{z'}_p}$; $(a_h)^{z'}_{i^{z'}_p} = (a_{h+1})^{z'}_{i^{z'}_p}$ при $h = \overline{h', m^{z'}_{i^{z'}_p} - 1}$;

– $m^{z'}_{i^{z'}_p} = m^{z'}_{i^{z'}_p} - 1$; если $m^{z'}_{i^{z'}_p} \neq 0$, то сформировано промежуточное решение $N^{z'}_T$ по составу группы $N^{z'}$, которое будет дополняться ПД $i^{q'}$-ых типов в количестве $r^{q'}$ элементов из множества Q; если $m^{z'}_{i^{z'}_p} = 0$, то $N^{z'}_T = N^{z'}_T \backslash \{[i^{z'}_p, m^{z'}_{i^{z'}_p}, (A)^{z'}_{i^{z'}_p}]\}$; $k_{z'} = k_{z'} - 1$;

7) для множества Q наборов параметров вида $Q = \{[i^q, m^q_i, (A)^q_i]_k | k = \overline{1, k_q}\}$ выполняется инициализация номера k рассматриваемого набора: $k = 1$;

8) определяется, в соответствии с номером k набора $[i^q, m^q_i, (A)^q_i]_k$, тип данных $i^{q'}$, пакеты которых будут размещаться в группе $N^{z'}$; для рассматриваемого типа данных $i^{q'}$ идентифицируется соответствующее ему множество $H_{i^{q'}}$;

9) если $H_{i^{q'}} \neq \emptyset$, то реализуется инициализация соответствующего множества $H^T_{i^{q'}}$: $H^T_{i^{q'}} = H_{i^{q'}}$; если $H_{i^{q'}} = \emptyset$, то выполняется переход на шаг 16;

10) для текущего значения параметра k в множестве Q осуществляется определение набора $N_2 = [i^{q'}, m^q_{i^{q'}}, (A)^q_{i^{q'}}] \in Q$, с которым выполняются преобразования;

11) проверка условия $H^T_{i^{q'}} \neq \emptyset$, в случае его истинности реализуется выделение в множестве $H^T_{i^{q'}}$ идентификатора $h^{q'}$ ПД $i^{q'}$-го типа, который будет размещаться в группе $N^{z'}$: $h^{q'} = \min\{h^T_{i^q} | h^T_{i^q} \in H^T_{i^q}\}$; модификация множества $H^T_{i^{q'}}$: $H^T_{i^{q'}} = H^T_{i^{q'}} \backslash \{h^{q'}\}$; для набора $N_2 = [i^{q'}, m^q_{i^{q'}}, (A)^q_{i^{q'}}] \in Q$ реализуется определение в векторе $(A)^q_{i^{q'}}$ в соответствии с индексом $h^{q'}$ количества данных $r^{q'}$ в пакете: $r^{q'} = (a_{h^q})^q_{i^{q'}}$; если условие $H^T_{i^{q'}} \neq \emptyset$ ложно, то реализуется переход на шаг 16;

12) проверка выполнения условия $i^{z'}_p = i^{q'}$, если условие не выполняется, то реализуется переход на шаг 13, при выполнении условия реализуется проверка условия $r^{z'}_p = r^{q'}$; если условие выполняется, то осуществляется переход на шаг 11, если условие не выполняется, то переход на шаг 13;

13) в решение $N^{z'}_T$ по составу группы $N^{z'}$ добавляется ПД $i^{q'}$-го типа из множества Q (итоговым решением является $N^{z'}_p$, предварительно выполнено присваивание $N^{z'}_p = N^{z'}_T$); для $i^{q'}$-го типа данных, пакет которых добавляется в группу $N^{z'}$, выполняются следующие действия:

если $(\exists k | i^{z'} = i^{q'}$ и $[i^{z'}, m^{z'}_i, (A)^{z'}_i]_k \in N^{z'}_p, k = \overline{1, k^p_{z'}})$, то $m^{z'}_i = m^{z'}_i + 1$; $(a_{m^{z'}_i})^{z'}_i = r^{q'}$;

если $(\overline{\exists} k | i^{z'} = i^{q'}$ и $[i^{z'}, m^{z'}_i, (A)^{z'}_i]_k \in N^{z'}_p, k = \overline{1, k^p_{z'}})$, то $i^{z'} = i^{q'}$; $m^{z'}_i = 1$; $(a_{m^{z'}_i})^{z'}_i = r^{q'}$; $k^p_{z'} = k^p_{z'} + 1$; $k = k^p_{z'}$; $N^{z'}_p = N^{z'}_p \cup \{[i^{z'}, m^{z'}_i, (A)^{z'}_i]_k\}$;

14) полученное промежуточное решение $N^{z'}_p$ передается на третий уровень иерархии для формирования расписания $\pi^{z'}$; выполняется проверка ограничения (1) на длительность обработки ПД, входящих в группу $N^{z'}$; если ограничение (1) не выполняется, то осуществляется переход на шаг 11; при выполнении условия (1) реализуется переход на шаг 15;

15) вычисление значения критерия f_2 для полученного решения по составам групп ПД $\{N^z | z = \overline{1, Z}\}$, в которое входит решение $N^{z'}_p$ по составу группы $N^{z'}$; для полученного значения f_2 реализуется вычисление левого $(\nabla^- f_2)$ либо правого $(\nabla^+ f_2)$ дискретных градиентов; если для сформированного решения $\{N^z | z = \overline{1, Z}\}$ с входящим в него решением $N^{z'}_p$ выполняется условие $(\nabla^- f_2) < 0$, то сравниваются значения $(\nabla^- f_2)$ и $(\nabla^- f_2)_p$ (с максимальным по модулю значением $(^- \nabla f_2)_p < 0$, полученным при формировании решений по составу группы $N^{z'}$); если условие $(\nabla^- f_2) < (\nabla^- f_2)_p$ выполняется, то реализуется сохранение типа данных $i^{q'}$ и количества данных $r^{q'}$ в пакете, добавление которого в группу $N^{z'}$ обеспечивает такое значение $(\nabla^- f_2)$, при котором $(\nabla^- f_2) < (\nabla^- f_2)_p$ (текущее лучшее решение в рамках окрестности решения $\{N^{z*} | z = \overline{1, Z}\}$ при формировании состава группы $N^{z'}$): $i^q_{w_p} = i^{q'}$; $r^q_{w_p} = r^{q'}$ (где $i^q_{w_p}$ – элемент с индексом w_p $(p = \overline{1,2})$ вектора $(i^q_w)_p$, предназначенного для хранения типов данных $i^{q'}$ в ПД, добавляемых в группу $N^{z'}$; $r^q_{w_p}$ – элемент с индексом w_p вектора $(r^q_w)_p$, предназначенного для хранения количества данных $r^{q'}$ в ПД, добавляемых в группу $N^{z'}$); выполняется модификация переменной-флага $p^{z'}$, указывающей на определение в процессе поиска более эффективного решения, чем текущее $\{N^{z*} | z = \overline{1, Z}\}$: $p^{z'} = 1$; реализуются присваивания $(\nabla^- f_2)_p = (\nabla^- f_2)$, $(\nabla^- f_2) = 0$; переход на шаг 11;

16) реализуется переход к рассмотрению ПД нового $i^{q'}$-го типа с целью добавления их в группу $N^{z'}$ за счет модификации индекса k набора вида $[i^{q'}, m^q_{i^{q'}}, (A)^q_{i^{q'}}]_k \in Q$: $k = k + 1$; если $k \leq k_q$, то выполняется переход на шаг 8; если $k > k_q$, то для группы $N^{z'}$ были сформированы все возможные решения, связанные с исключением из нее ПД $i^{z'}_p$-го

типа в количестве $r_p^{z'}$ элементов и добавлении в нее ПД i^q-ых типов, таких, что $[i^q, m_i^q, (A)_i^q]_k \in Q, k = \overline{1, k_q}$;

17) реализуется проверка условия, которое характеризует наличие лучшего решения в окрестности текущего эффективного решения $\{N^{z*} | z = \overline{1, Z}\}$ (при $p^{z'} = 1$); в случае, если данное условие не выполняется, то осуществляется переход на шаг 19; при истинности условия $p^{z'} = 1$ с использованием значений параметров $i^{q'}$ и $r^{q'}$ (для хранения использованы вектора $(i_w^q)_p$ и $(r_w^q)_p$, $w = \overline{1, w_p}$) формируется промежуточное локально оптимальное решение $N_T^{z'}$ по составу группы $N^{z'}$: для текущего значения w_p количества ПД, добавляемых в группу $N^{z'}$ (w_p– идентификатор последнего элемента в векторах $(i_w^q)_p$ и $(r_w^q)_p$), выполняется инициализация $i^{q'}$-го типа данных, ПД которых добавляется в группу $N^{z'}$ при формировании решения $N_T^{z'}$, и количества данных $r^{q'}$следующим образом: $i^{q'} = i_{w_p}^q, r^{q'} = r_{w_p}^q$, где $i_{w_p}^q$ и $r_{w_p}^q$ – последние элементы в векторах (i_w^q) и (r_w^q); формирование промежуточного решения $N_T^{z'}$ следующим образом:

если $(\exists k |\ i^{z'} = i^{q'}$ и $[i^{z'}, m_i^{z'}, (A)_i^{z'}]_k \in N_T^{z'}, k = \overline{1, k_{z'}})$, то $m_i^{z'} = m_i^{z'} + 1$; $(a_{m_i^{z'}})_i^{z'} = r^{q'}$;

если $(\overline{\exists} k |\ i^{z'} = i^{q'}$ и $[i^{z'}, m_i^{z'}, (A)_i^{z'}]_k \in N_T^{z'}, k = \overline{1, k_{z'}})$, тогда $i^{z'} = i^{q'}$; $m_i^{z'} = 1$; $(a_{m_i^{z'}})_i^{z'} = r^{q'}$; $k_{z'} = k_{z'} + 1$; $k = k_{z'}$; $N_T^{z'} = N_T^{z'} \cup \{[i^{z'}, m_i^{z'}, (A)_i^{z'}]_k\}$;

18) для промежуточного локально оптимального решения $N_T^{z'}$ выполняется проверка условия (1) добавления ПД в группу $N^{z'}$; при его выполнении возможно дополнение группы $N^{z'}$ новыми ПД, принадлежащими множеству Q (модификация промежуточного эффективного решения $N_T^{z'}$); ПД $i_{w_p}^q$-го типа в количестве $r_{w_p}^q$ элементов размещен в решении $N_T^{z'}$, и его идентификатор исключается из множества $H_{i^q}(i^q = i_{w_p}^q)$; для этого в множестве Q идентифицируется набор параметров $[i^q, m_i^q, (A)_i^q]_k \in Q$, для которого $i^q = i_{w_p}^q$, в соответствующем (i^q)-му типу данных $(i^q = i_{w_p}^q)$ векторе $(A)_i^q$ определяется индекс (номер) h элемента $(a_h)_i^q$, для которого $(a_h)_i^q = r_{w_p}^q$; определенный таким образом номер ПД i^q-го типа $(i^q = i_{w_p}^q)$ исключается из множества H_{i^q}: $H_{i^q} = H_{i^q} \backslash \{h\}$; увеличение номера w_p последнего элемента в векторах $(i_w^q)_p$ и $(r_w^q)_p$: $w_p = w_p + 1$; выполняется инициализация $p^{z'} = 0$ и $(\nabla^- f_2)_p = 0$; выполняется переход на шаг 7; если условие (1) не выполняется, то реализуется переход на шаг 19;

19) осуществляется изменение индекса p условия, с использованием которого определяется ПД, исключаемый из рассматриваемой группы $N^{z'}$: $p = p + 1$; если $p \le 2$, то выполняется инициализация номера w_p для индекса p: $w_p = 1$; переменная-флаг $p^{z'} = 0$; осуществляется переход на шаг 4; при $p > 2$ осуществляется переход на шаг 20 ;

20) сравнение значения $(\nabla^- f_2)_G$ со значениями $(\nabla^- f_2)_p$ $(p = \overline{1,2})$, характеризующими наилучшие варианты решений $\{N^z | z = \overline{1, Z}\}$, включающих решения $N_p^{z'}$ по составу рассматриваемой группы $N^{z'}$; если $(\nabla^- f_2)_1 < (\nabla^- f_2)_2$ и $(\nabla^- f_2)_1 < (\nabla^- f_2)_G$, то осуществляется сохранение типов данных и количества данных в пакетах, обмен которыми между группой $N^{z'}$ и множеством Q позволил получить значение $(\nabla^- f_2)$, равное $(\nabla^- f_2)_1$, сохранение параметров выполняется следующим образом:

– $z1 = z', i^z = i_1^{z'}; r^z = r_1^{z'}$ (параметры ПД, извлекаемого из группы $N^{z'}$);

– $w_G = w_1$, $(i_w^q)_G = (i_w^q)_1$; $(r_w^q)_G = (r_w^q)_1$ (параметры добавляемых в группу $N^{z'}$ пакетов, если $w = \overline{1, w_1}$);

– $(\nabla^- f_2)_G = (\nabla^- f_2)_1$ (значение $(\nabla^- f_2)_G$ сохранено); переход на шаг 22;

21) в случае, если при выполнении условия $(\nabla^- f_2)_1 < (\nabla^- f_2)_2$ условие $(\nabla^- f_2)_1 < (\nabla^- f_2)_G$ не выполняется, то реализуется переход на шаг 22; если условие $(\nabla^- f_2)_1 < (\nabla^- f_2)_2$ не выполняется (условие вида $(\nabla^- f_2)_2 \le (\nabla^- f_2)_1$ является истинным), то осуществляется проверка условия $(\nabla^- f_2)_2 < (\nabla^- f_2)_G$; в случае истинности условия $(\nabla^- f_2)_2 < (\nabla^- f_2)_G$ выполняется сохранение решения по составу группы $N^{z'}$:

– $z1 = z'; i^z = i_2^{z'}; r^z = r_2^{z'}$ (параметры извлекаемого из группы $N^{z'}$ ПД);

– $w_G = w_2$, $(i_w^q)_G = (i_w^q)_2$; $(r_w^q)_G = (r_w^q)_2$ (параметры добавляемых в группу $N^{z'}$ пакетов);

– $(\nabla^- f_2)_G = (\nabla^- f_2)_2$, реализуется переход на шаг 22; если условие $(\nabla^- f_2)_2 \le (\nabla^- f_2)_1$ выполняется, а условие $(\nabla^- f_2)_2 < (\nabla^- f_2)_G$ – нет, то реализуется переход на шаг 22;

22) изменение номера группы $N^{z'}$, состав которой будет модифицирован: $z' = z' + 1$; проверка условия $z' \le Z$, если условие является истинным, то реализуется переход на шаг 3; в случае не выполнения условия – переход на шаг 23;

23) реализуется проверка условия $(\nabla^- f_2)_G < 0$, если условие выполняется, то в окрестности текущего локально оптимального решения $\{N^{z*} | z = \overline{1, Z}\}$ получено лучшее решение, значения параметров которого сохранены в переменных $z1, i^z$, r^z– параметры ПД, извлекаемого из группы N^z, $(i_w^q)_G, (r_w^q)_G$ (при $w = \overline{1, w_G}$) – параметры ПД, добавляемых в группу $N^{z'}$; реализуется формирование нового локально оптимального решения $\{N^{z*} | z = \overline{1, Z}\}$, для этого реализуется действия:

– инициализация номера группы ПД $z = z1$; по типу данных i^z в группе N^z, как в множестве наборов параметров $N^z = \{[i^z, m_i^z, (A)_i^z]_k | k = \overline{1, k_z}\}$, проводится идентификация набора, с которым будут выполняться преобразования, $N_1 = [i^z, m_{i^z}^z, (A)_{i^z}^z] \in N^z$, с которым будут выполняться действия; в исход-

ном решении по составу группы N^z, входящем в решение $\{N^{z*} | z = \overline{1,Z}\}$), реализуется исключение ПД i^z-го типа следующим образом:

для i^z-го типа данных и их количества r^z (в ПД, исключаемом из группы N^z) определение для набора $N_1 = [i^z, m_{i^z}^z, (A)_{i^z}^z] \in N^z$ в векторе $(A)_{i^z}^z$ индекса h' пакета такого, что $(a_{h'})_{i^z}^z = r^z$ (при $z = z1$);

переупорядочивание элементов вектора $(A)_{i^z}^z$; $(a_h)_{i^z}^z = (a_{h+1})_{i^z}^z$ при $h = \overline{h', m_{i^z}^z - 1}$; $m_{i^z}^z = m_{i^z}^z - 1$;

если $m_i^z \neq 0$, то сформировано промежуточное решение по составу группы N^z (будут дополняться ПД i_w^q-ых типов в количестве r_w^q элементов множества Q); если $m_{i^z}^z = 0$, то $N^z = N^z \backslash \{[i^z, m_{i^z}^z, (A)_{i^z}^z]\}$; $k_z = k_z - 1$;

– дополнение множества Q ПД i^z-го типа в количестве r^z элементов; действия по размещению ПД i^z-го типа в множестве Q реализуются таким образом:

если $(\exists\, k |\ i^q = i^z$ при $[i^q, m_i^q, (A)_i^q]_k \in Q, k = \overline{1,k_q})$, то $m_i^q = m_i^q + 1$; $(a_{m_i^q})_i^q = r^z$;

если $(\overline{\exists}\, k |\ i^q = i^z$ при $[i^q, m_i^q, (A)_i^q]_k \in Q, k = \overline{1,k_q})$, то $i^q = i^z$; $m_i^q = 1$; $(a_{m_i^q})_i^q = r^z$; $k_q = k_q + 1$; $k = k_q$; $Q = Q \cup \{[i^q, m_i^q, (A)_i^q]_k\}$;

– задание индекса w в массивах $(i_w^q)_p$ и $(r_w^q)_p$ для ПД, добавляемых в группу N^z, равным 1 ($w = 1$) (значение w_p количества ПД, добавляемых в группу N^z, определено в процессе идентификации локально оптимального решения в окрестности $\{N^{z*} | z = \overline{1,Z}\}$);

– инициализация $i^{q'}$-го типа данных, пакет которых добавляется в группу N^z, определение количества данных $r^{q'}$ в добавляемом ПД следующим образом: $i^{q'} = i_w^q, r^{q'} = r_w^q$,

– исключение ПД $i^{q'}$-го типа в количестве $r^{q'}$ элементов из множества Q, которое реализуется следующим образом:

для $i^{q'}$-го типа данных и их количества $r^{q'}$ определение набора параметров $N_2 = [i^{q'}, m_i^{q'}, (A)_i^{q'}] \in Q$, в векторе $(A)_i^{q'}$ которого идентифицируется индекс h' ПД такой, что $(a_{h'})_i^{q'} = r^{q'}$;

переупорядочивание элементов вектора $(A)_i^{q'}$; $(a_h)_i^{q'} = (a_{h+1})_i^{q'}$ при $h = \overline{h', m_i^{q'} - 1}$;

$m_i^{q'} = m_i^{q'} - 1$; если $m_i^{q'} \neq 0$, то удаление ПД $i^{q'}$-го типа завершено; $m_i^{q'} = 0$, $Q = Q \backslash \{[i^{q'}, m_i^{q'}, (A)_i^{q'}]\}$; $k_q = k_q - 1$;

– добавление ПД $i^{q'}$-го типа в количестве $r^{q'}$ элементов в группу N^z осуществляется следующим образом:

если $(\exists\, k |\ i^z = i^{q'}$ при $[i^z, m_i^z, (A)_i^z]_k \in N^z, k = \overline{1,k_z})$, то $m_i^z = m_i^z + 1$; $(a_{m_i^z})_i^z = r^{q'}$;

если $(\overline{\exists}\, k |\ i^z = i^{q'}$ при $[i^z, m_i^z, (A)_i^z]_k \in N^z, k = \overline{1,k_z})$, тогда $i^z = i^{q'}$; $m_i^z = 1$; $(a_{m_i^z})_i^z = r^{q'}$; $k_z = k_z + 1$; $k = k_z$; $N^z = N^z \cup \{[i^z, m_{i^z}^z, (A)_{i^z}^z]_k\}$;

– для векторов $(i_w^q)_p$ и $(r_w^q)_p$ параметров ПД, добавляемых в группу N^z, изменение значения индекса w: $w = w + 1$; если $w \leq w_p$, то реализуется переход к добавлению новых пакетов в N^z; если выполняется условие $w > w_p$, то эффективное решение по составу группы N^z сформировано (в итоге сформировано новое локально оптимальное решение по составу групп ПД $\{N^{z*} | z = \overline{1,Z}\}$); осуществляется переход на шаг 1; если условие $(\nabla^- f_2)_G < 0$ не является истинным, то в окрестности решения $\{N^{z*} | z = \overline{1,Z}\}$ не найдено более эффективное решение, выполняется переход на шаг 24;

24) останов алгоритма.

Вычислительная сложность алгоритма определена как $O(n \cdot n^i \cdot Z)$. В качестве параметра, характеризующего эффективность применения метода оптимизации групп ПД, определялось отношение разности значений общих простоев приборов КС (в течение интервалов t^z ($z = \overline{1,Z}$)) при формировании фиксированных групп пакетов (полученных при построении начального решения) и групп пакетов, составы которых сформированы в соответствии с рассмотренным методом, к значению общих простоев приборов КС для фиксированных групп N^z: $f_{\text{эмосгп}} = (f_2^{\text{фикс}} - f_2^{\text{мосгп}})/f_2^{\text{фикс}}$, где $f_{\text{эмосгп}}$ – параметр, характеризующий степень снижения простоев приборов КС в течение всех интервалов времени ее функционировании; $f_2^{\text{фикс}}$ – значение, характеризующее простои приборов КС с течением интервалов времени ее функционирования для решений по составом фиксированных групп ПЗ (начальные решения по составам групп ПЗ); $f_2^{\text{мосгп}}$ – значение, характеризующее простои приборов КС в течением интервалов времени ее функционирования для решений по составом групп ПЗ, оптимизированных с использованием предложенного метода.

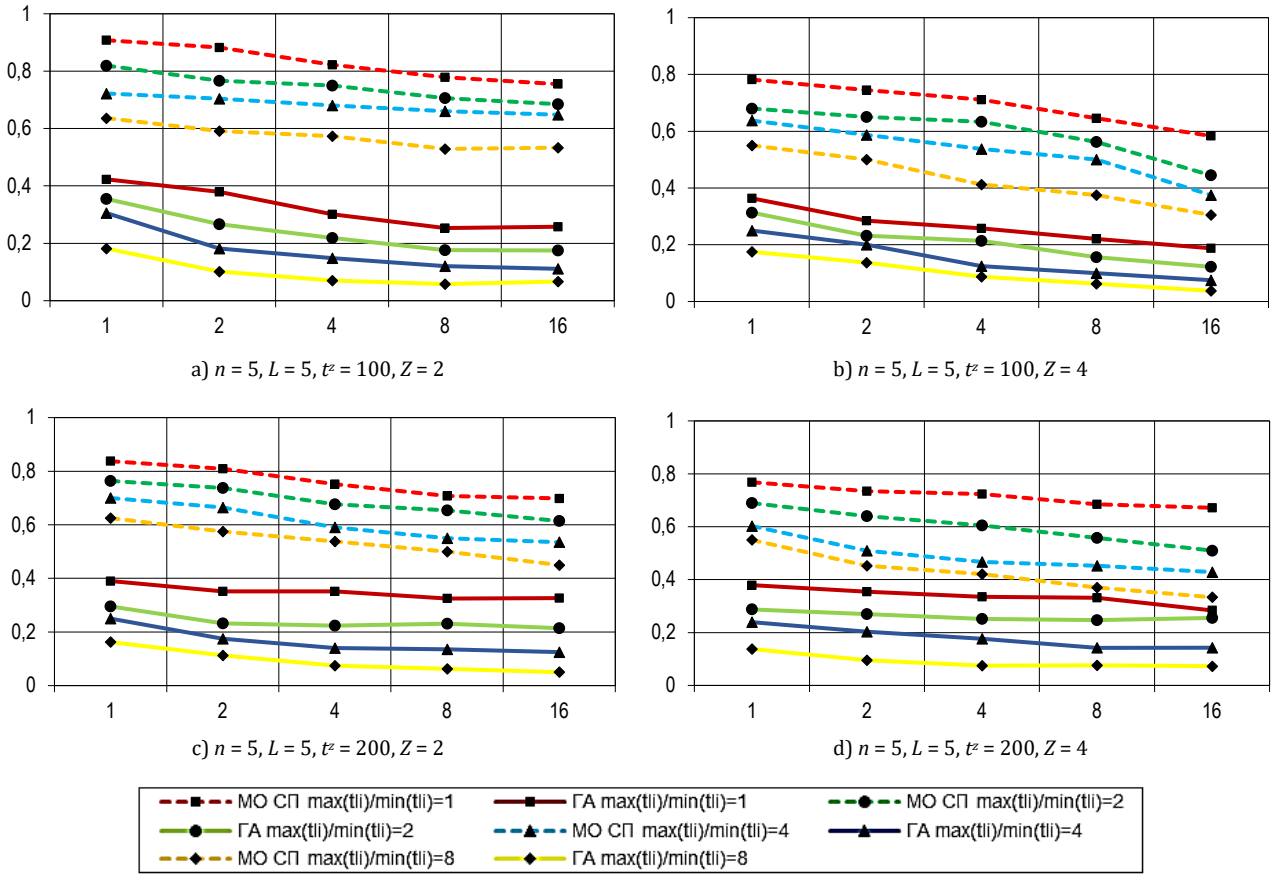
4. Анализ результатов исследований эффективности метода определения составов групп ПД и их сравнение с результатами, полученными метаэвристическими алгоритмами

С целью исследования эффективности метода определения составов групп ПД в рассмотрение введены следующие обозначения [24]: $\max(t_i^l)/\min(t_i^l)$ – отношение максимальной длительности обработки данных i-ых типов на l-ых сегментах конвейера к минимальной длительности обработки данных; $\max(t_{ij}^l)/\min(t_{ij}^l)$ – отношение максимальной длительности переналадки сегментов конвейера ($l = \overline{1,L}$) к минимальной длительности переналадки сегментов. Задаваемые при исследовании значения параметров: $n = 5$; $L = 5$; $n^i = 24$.

Значения параметра t^z задавались равными 100 и 200, значения параметра Z – соответственно, 2 и 4. Значения $\max(t_{ij}^l)/\min(t_{ij}^l)$ задавались равными 1, 2, 4, 8, 16, а значения отношения $\max(t_i^l)/\min(t_i^l)$ – равными 1, 2, 4, 8. При исследованиях эффективности метода определения групп ПД формировались следующие решения: фиксированные группы ПД (начальные решения по составам групп ПД, предполагающие распределение пакетов из решения [M,A] с верхнего уровня по группам $N^z(z = \overline{1,Z})$ без их дальнейшей оптимизации); оптимизированные составы групп ПД (с использованием рассмотренного метода на основе начальных решений по составам групп $N^z(z = \overline{1,Z})$ сформированы эффективные решения $N^{z*}(z = \overline{1,Z})$).

Эффективность метода формирования групп ПД сравнивается с эффективностью применения метаэвристических алгоритмов (ГА). Метод оптимизации составов групп ПД, использующий ГА, предполагает, что в каждой хромосоме, соответствующей решению, выделяется n участков [24]. Каждый участок соответствует определенному i-му типу данных, количество генов в каждом участке равно значению m_i для этого типа данных в решении $[M, A]$. Значение q_{h_i} гена в i-ом участке хромосомы, соответствующего h_i-му ПД ($h_i = \overline{1,m}$), характеризует принадлежность этого ПД одной из групп $N^z(z = \overline{1,Z})$, либо (при $q_{h_i} = 0$) – принадлежность множеству Q ($q_{h_i} \in \{0, 1, .., Z\}$). Формирование начальной популяции предполагает генерацию значений q_{h_i} случайным образом и проверку выполнения для сформированных решений по составам групп ПД условия (1) с учетом построенного для них расписания. Скрещивание хромосом выполняется в точках, соответствующих границам участков для i-ых типов данных. Мутация генов предусматривает изменение текущих их значений с учетом условия допустимости сформированных групп.

Графики зависимостей эффективности применения метода оптимизации групп ПД от значений входных параметров (n = 5, L = 5, $t^z = 100$ и t^z = 200, Z = 2 и Z = 4, при изменении значений $\max(t_{ij}^l)/\min(t_{ij}^l)$ и $\max(t_i^l)/\min(t_i^l)$) в сравнении с применением ГА представлены на рисунке 1, где МО СГП – метод оптимизации составов групп пакетов. По оси ординат откладывается степень снижения общих простоев приборов конвейера при обработки ПД всех групп для решения, полученного после оптимизации их составов, по сравнению с начальным решением по распределению ПД по группам (значения $f_{\text{эмосгп}}$), а по оси абсцисс – значения $\max(t_{ij}^l)/\min(t_{ij}^l)$.



a) n = 5, L = 5, t^z = 100, Z = 2

b) n = 5, L = 5, t^z = 100, Z = 4

c) n = 5, L = 5, t^z = 200, Z = 2

d) n = 5, L = 5, t^z = 200, Z = 4

Рис. 1. Снижение общих простоев сегментов при оптимизации составов групп ПЗ

Fig. 1. Decrease of Total Segment Downtime When Optimizing the Composition of the Data Packets Groups

Анализ результатов исследований метода оптимизации составов групп ПД позволил сформулировать следующие особенности формирования решений в соответствии с введенными в рассмотрение значениями параметров:

– при малых значениях отношений $\max(t_i^l)/\min(t_i^l)$ и $\max(t_{ij}^l)/\min(t_{ij}^l)$ и (равных 1 и 2) метод оптимизации составов групп ПЗ позволяет получить решения, которые обеспечивают наиболее эффективное использование ограниченного ресурса времени функционирования КС; т. е. незначительная неоднородность длительностей обработки данных и длительностей переналадки приборов на обработку данных соответствующих типов позволяет рассматриваемому методу сформировать группы ПЗ, которые обеспечивают в полной мере использование временного ресурса приборов КС;

– по мере роста неоднородностей длительностей переналадок и неоднородностей длительностей обработки данных степень снижения суммарных простоев приборов в решениях, полученных с использованием рассмотренного метода, по сравнению с решениями, соответствующими фиксированным группам пакетов, уменьшается (уменьшение на 25–30 % при увеличении значений отношения $\max(t_i^l)/\min(t_i^l)$ от 1 до 8, уменьшение на 15–25 % при увеличении значений отношения $\max(t_{ij}^l)/\min(t_{ij}^l)$ от 1 до 16); это связано с тем, что увеличение неоднородностей указанных видов не позволяет сформировать составы групп N^z, в которых обработка ПД обеспечивает наилучшее использованием ограниченного временно́го ресурса КС (значительные простои сегментов конвейера в ожидании готовности к выполнению заданий в пакетах и простои приборов в ожидании готовности ПД к выполнению после их (приборов) переналадки);

– увеличение количества групп ПД (количество рассматриваемых интервалов времени t^z) обуславливает незначительное (10–15 %) уменьшение значения $f_{\text{эмосгп}}$ (накапливается (суммируется) общая неэффективность использования ограниченного временного ресурса КС);

– увеличение длительностей интервалов функционирования КС обуславливает уменьшение значений $f_{\text{эмосгп}}$ на 5–8 %; это связано с тем, что при увеличении количества интервалов большее количество ПД может быть размещено в группах для их выполнения (т. е. при $t^z = 200$ группы содержат большее количество ПД, чем при $t^z = 100$); в силу наличия неоднородности длительностей обработки данных на разных приборах и неоднородности длительностей переналадок решения по составам ПД не могут гарантировать лучшего использования ресурса времени КС.

Таким образом, анализ результатов исследований при $n = 5$, $L = 5, t^z = 100$ и $t^z = 200$, $Z = 2$ и $Z = 4$, при изменении значений параметров $\max(t_{ij}^l)/\min(t_{ij}^l)$ и $\max(t_i^l)/\min(t_i^l)$ в указанных диапазонах показал, что максимальное снижение суммарных простоев приборов в решениях, сформированных с использованием рассмотренного метода, по сравнению с решениями, соответствующими фиксированным группам пакетов, получено при значениях $t^z = 100$, $Z = 2$, при $\max(t_i^l)/\min(t_i^l) = 1$, $\max(t_{ij}^l)/\min(t_{ij}^l) = 1$ и составляет в среднем 65 %. Т. е. при минимальных неоднородностях длительностей обработки данных разных типов на приборах КС и длительностей переналадок приборов с обработки данных одного типа на обработку данных другого типа степень снижения суммарных простоев приборов в решениях, полученных с использованием рассмотренного метода, по сравнению с решениями, соответствующими фиксированным группам пакетов, является максимальной. Степень снижения суммарных простоев приборов в решениях, полученных с использованием рассмотренного метода, по сравнению с решениями, соответствующими фиксированным группам пакетов, уменьшается при увеличении значений отношений $\max(t_i^l)/\min(t_i^l)$ и $\max(t_{ij}^l)/\min(t_{ij}^l)$. При этом увеличение неоднородности длительностей обработки данных разных типов на сегментах конвейера влияет на уменьшение значения $f_{\text{эмосгп}}$ более значительно, чем увеличение неоднородностей длительностей переналадок этих сегментов. Увеличение количества групп ПД (с $Z = 2$ до $Z = 4$) при прочих равных значениях входных параметров приводит к уменьшению степени снижения суммарных простоев приборов в решениях, полученных с использованием рассмотренного метода, по сравнению с решениями, соответствующими фиксированным группам пакетов, на 10 % и более. Увеличение значений длительностей интервалов функционирования КС с $t^z = 100$ до $t^z = 200$ также приводит к уменьшению степени снижения суммарных простоев приборов в решениях, полученных с использованием рассмотренного метода, по сравнению с решениями, соответствующими фиксированным группам пакетов, на 5-7 %. Сравнение результатов применения разработанного метода оптимизации составов групп ПД с применением ГА показало, что использование предложенного метода позволяет получить решения, которые на 45 % более эффективны, чем результаты, полученные с использованием ГА.

Заключение

Разработан иерархический подход к комплексному планированию обработки ПД при наличии ограничений и метод оптимизации составов групп ПД, обрабатываемых в течение заданных временных интервалов. Метод оптимизации составов групп $N^z(z = \overline{1, Z})$ предусматривает определение таких ПД, обработка которых в их позициях в по-

следовательностях π^l $(l = \overline{1,L})$ расписания обуславливает простои сегментов конвейера, являющиеся наибольшими среди всех пакетов, находящихся в этих последовательностях. Исключение пакетов из состава групп $N^z(z = \overline{1,Z})$, размещение в группах ПД, не включенных в группы, позволяет сформировать окрестности текущих локально оптимальных решений, в которых идентифицируются решения по составам групп ПД. Использование рассмотренного метода позволяет в среднем до 65 % уменьшить простои сегментов конвейера по сравнению с составами групп, не предусматривающими оптимизацию и получить решения, которые на 45 % являются более эффективными, чем решения, полученные с использованием ГА.

Список используемых источников

1. Кротов К.В. Комплексный метод определения эффективных решений по составам партий данных и расписаниям их обработки в конвейерных системах // Вычислительные технологии. 2018. Т. 23. № 3. С. 58–76.
2. Mendez C.A., Cerdá J., Grossmann I.E., Harjunkoski I., Fahl M. State-of-the-art review of optimization methods for short-term scheduling of batch processes // Computers and Chemical Engineering. 2006. Vol. 30. Iss. 6–7. PP. 913–946. DOI:10.1016/j.compchemeng.2006.02.008
3. Adonyi R., Romero J., Puigjaner L., Friedler F. Incorporating heat integration in batch process scheduling // Applied Thermal Engineering. 2003. Vol. 23. Iss. 14. PP. 1743–1762. DOI:10.1016/S1359-4311(03)00141-8
4. Agha M. Integrated Management of Energy and Production: Scheduling of Batch Process and Combined Heat & Power Plant. PhD Thesis. Toulouse, France: University of Toulouse, 2009. 255 p.
5. Zeballos L.J., Henning G.P. A CP approach to the scheduling of resource-constrained multiproduct continuous facilities // Latin American Applied Research. 2006. Iss. 36. PP. 205–212.
6. Díaz-Ramírez J., Huertas J.I. A continuous time model for a short-term multiproduct batch process scheduling // Ingeniería e Investigación. 2018. Vol. 38. Iss. 1. PP. 96–104. DOI:10.15446/ing.investig.v38n1.66425
7. Ковалев М.Я. Модели и методы календарного планирования. Минск: БГУ, 2004. URL: https://www.studmed.ru/kovalev-mya-modeli-i-metody-kalendarnogo-planirovaniya_ca57fbd.html (дата обращения 10.09.2020)
8. Chandra P., Gupta S. Managing batch processors to reduce lead time in a semiconductor packaging line // International Jornal of Production Research. 1997. Vol. 35. Iss. 3. PP. 611–633. DOI:10.1080/002075497195623
9. Bein W., Noga J., Wiegley J. Priority Approximation for Batching // Proceedings of the 41st Hawaii International Conference on Systems Science (HICSS-41 2008, Waikoloa, Big Island, Hawaii, 7–10 January 2008). IEEE, 2008. DOI:10.1109/HICSS.2008.508
10. Bein W., Noga J., Wiegley J. Approximation for Batching via Priorities // Scientific Annals of Computer Science. 2017. Vol. 17. PP. 1–18.
11. Steiner G., Zhang R. Minimizing the weight number of late jobs with Batch setup times and delivery costs on a single machine. Vienna: Itech Education and Published, 2007. DOI:10.5772/5216
12. Koehler F. Khuller S. Optimal Batch Schedules for Parrallel Machines // Proceedings of the 13th International Symposium on Workshop on Algorithms and Data Structures (WADS 2013, London, ON, Canada, 12–14 August 2013). Lecture Notes in Computer Science. Berlin: Springer, 2013. Vol. 8037. PP. 475–486. DOI:10.1007/978-3-642-40104-6_41
13. Monch L., Balasubramanian H., Fowler J. W., Pfund M. E. Heuristic scheduling of jobs on parallel batch machines with incompatible job families and unequal ready times // Computers & Operations Research. 2005. Vol. 32. Iss. 11. PP. 2731–2750. DOI:10.1016/j.cor.2004.04.001
14. Dang Th.T., Frankovic B., Budinska I., Flood B., Sheahan C., Dang B.L. Using heuristic search for solving single machine batch processing problems // Computing and Informatics. 2006. Vol. 25. Iss. 5. PP. 405–420.
15. Kohn R., Rose O., Laroque Ch. Study on multi-objective optimization for parallel batch machine scheduling using variable neighbourhood search // Proceedings of the Winter Simulation Conference (WSC, Washington, USA, 8–11 December 2013). IEEE, 2013. PP. 3654–3670. DOI:10.1109/WSC.2013.6721726
16. Cheng T., Yuan J., Yang A. Scheduling a batch-processing machine subject to precedence constraints, release dates and identical processing times // Computers and Operations Research. 2005. Vol. 32. Iss. 4. PP. 849–859. DOI:10.1016/j.cor.2003.09.001
17. Van der Zee D.-J. Dynamic scheduling of batch processing machines with non-identical product sizes // International Journal of Production Research. 2007. Vol. 45. Iss. 10. PP. 2327–2349. DOI:10.1080/00207540600690537
18. Tan Y., Huangi W., Sun Y., Yue Y. Comparative study of different approaches to solve batch process sheduling and optimisation problems // Proceedings of the 18th International Conference on Automation and Computing (ICAC, Loughborough, UK, 7–8 September 2012). IEEE, 2012. PP. 424–444.
19. Cheng B.-Y., Chen H.-P., Wang S.-S. Improved Ant Colony Optimization Method for Single Batch-Processing Machine with Non-Identical Job Sizes // Journal of System Simulation. 2009. Vol. 21. Iss. 9. PP. 2687–2695.
20. Ramasubramanian M., Mathirajan M. A Mathematical Model for Scheduling a Batch Processing Machine with Multiple Incompatible Job Families, Non-identical Job dimensions, Nonidentical Job sizes, Non-agreeable release times and due dates // Proceedings of the International Conference on Manufacturing, Optimization, Industrial and Material Engineering (MOIME 2013, Grand Hyatt, Bandung, Indonesia, 9–10 March 2013). IOP Conference Series: Materials Science and Engineering. 2013. Vol. 46. DOI:10.1088/1757-899X/46/1/012013
21. Ogun B., Çigdem A.-U. Mathematical models for a batch scheduling problem to minimize earliness and tardiness // Journal of Industrial Engineering and Management. 2018. Vol. 11. Iss. 3. PP. 390–405. DOI:10.3926/jiem.2541
22. Kreipl S., Pinedo M. Planning and Scheduling in Supply Chains: An Overview of Issues in Practice // Production and Operations Management. 2004. Vol. 13. Iss. 1. PP. 77–92. DOI:10.1111/j.1937-5956.2004.tb00146.x

23. Кротов К.В. Обоснование модели многоуровневого программирования для построения расписаний групповой обработки партий данных в конвейерной системе при наличии ограничений // Вестник Иркутского государственного технического университета. 2016. №1(108). С. 35–48.

24. Кротов К.В. Использование генетических алгоритмов для построения эффективных комплексных расписаний обработки пакетов данных в конвейерной системе при задании ограничений на длительность интервалов времени ее функционирования // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. 2020. № 1. С. 71–86. DOI:10.17308/sait.2020.1/2594

* * *

Building Complex Schedules of Data Packets Processing with Setting Time Limits of a Conveyor System Functioning

K. Krotov[1]

[1]Sebastopol State University,
Sebastopol, 299053, Russian Federation

Article info
DOI:10.31854/1813-324X-2020-6-3-75-90
Received 18th February 2020
Accepted 1st September 2020

For citation: Krotov K. Building Complex Schedules of Data Packets Processing with Setting Time Limits of a Conveyor System Functioning. *Proc. of Telecom. Universities.* 2020;6(3):75–90. (in Russ.) DOI:10.31854/1813-324X-2020-6-3-75-90

Abstract: *The problem of planning data packet processing in a pipeline system with a time limit on the duration of its operation intervals is considered. The solution of the problem involves determining the composition of data packets, the composition of groups of data packets processed during these time intervals, and the schedules for processing packets of each group. To optimize solutions, the hierarchical game theory is applied. Conditions have been introduced that allow you to determine packages that are processed or readjusted to processing, which causes maximum downtime of pipeline segments. A method for constructing effective group compositions is proposed, which involves excluding packages that are determined in accordance with these conditions and placing packages that are not included in them in groups.*

Keywords: *hierarchical game, intervals of conveyor system functioning, data packets groups, local optimization.*

References

1. Krotov K.V. The Complex Method for Definition of Effective Decisions on Data Set Composition and Shedule for Their Processing in Conveyor Systems. *Computational technologies*. 2018;23(3):58–76. (in Russ.)

2. Mendez C.A., Cerdá J., Grossmann I.E., Harjunkoski I., Fahl M. State-of-the-art review of optimization methods for short-term scheduling of batch processes. *Computers and Chemical Engineering*. 2006;30(6-7):913–946. DOI:10.1016/j.compchemeng.2006.02.008

3. Adonyi R., Romero J., Puigjaner L., Friedler F. Incorporating heat integration in batch process scheduling. *Applied Thermal Engineering*. 2003;23(14):1743–1762. DOI:10.1016/S1359-4311(03)00141-8

4. Agha M. *Integrated Management of Energy and Production: Scheduling of Batch Process and Combined Heat& Power Plant*. PhD Thesis. Toulouse, France: University of Toulouse; 2009. 255 p.

5. Zeballos L.J., Henning G.P. A CP approach to the scheduling of resource-constrained multiproduct continuous facilities. *Latin American Applied Research*. 2006;36:205–212.

6. Díaz-Ramírez J., Huertas J.I. A continuous time model for a short-term multiproduct batch process scheduling. *Ingeniería e Investigación*. 2018;38(1):96–104. DOI:10.15446/ing.investig.v38n1.66425

7. Kovalev M.Ya. *Scheduling Models and Methods*. Minsk: Belarusian State University Publ.; 2004. Available from: https://www.studmed.ru/kovalev-mya-modeli-i-metody-kalendarnogo-planirovaniya_ca57fbd.html [Accessed 10th September 2020]

8. Chandra P., Gupta S. Managing batch processors to reduce lead time in a semiconductor packaging line. *International Jornal of Production Research*. 1997;35(3):611–633. DOI:10.1080/002075497195623

9. Bein W., Noga J., Wiegley J. Priority Approximation for Batching. *Proceedings of the 41st Hawaii International Conference on Systems Science, HICSS-41 2008, 7–10 January 2008, Waikoloa, Big Island, Hawaii*. IEEE; 2008. DOI:10.1109/HICSS.2008.508

10. Bein W., Noga J., Wiegley J. Approximation for Batching via Priorities. *Scientific Annals of Computer Science*. 2017;17:1–18.

11. Steiner G., Zhang R. *Minimizing the weight number of late jobs with Batch setup times and delivery costs on a single machine*. Vienna: Itech Education and Published; 2007. 436 p. DOI:10.5772/5216

12. Koehler F. Khuller S. Optimal Batch Schedules for Parrallel Machines. *Proceedings of the 13th International Symposium on Workshop on Algorithms and Data Structures, WADS 2013, 12–14 August 2013, London, ON, Canada. Lecture Notes in Computer Science*. Berlin: Springer; 2013. vol.8037. p.475–486. DOI:10.1007/978-3-642-40104-6_41

13. Monch L., Balasubramanian H., Fowler J. W., Pfund M. E. Heuristic scheduling of jobs on parallel batch machines with incompatible job families and unequal ready times. *Computers & Operations Research*. 2005;32(11):2731–2750. DOI:10.1016/j.cor.2004.04.001

14. Dang Th.T., Frankovic B., Budinska I., Flood B., Sheahan C., Dang B.L. Using heuristic search for solving single machine batch processing problems. *Computing and Informatics*. 2006;25(5):405–420.

15. Kohn R., Rose O., Laroque Ch. Study on multi-objective optimization for parallel batch machine scheduling using variable neighbourhood search. *Proceedings of the Winter Simulation Conference, WSC, 8–11 December 2013, Washington, USA*. IEEE; 2013. p.3654–3670. DOI:10.1109/WSC.2013.6721726

16. Cheng T., Yuan J., Yang A. Scheduling a batch-processing machine subject to precedence constraints, release dates and identical processing times. *Computers and Operations Research*. 2005;32(4):849–859. DOI:10.1016/j.cor.2003.09.001

17. Van der Zee D.-J. Dynamic scheduling of batch processing machines with non-identical product sizes. *International Journal of Production Research*. 2007;45(10):2327–2349. DOI:10.1080/00207540600690537

18. Tan Y., Huangi W., Sun Y., Yue Y. Comparative study of different approaches to solve batch process sheduling and optimisation problems. *Proceedings of the 18th International Conference on Automation and Computing, ICAC, 7–8 September 2012, Loughborough, UK*. IEEE; 2012. p.424–444.

19. Cheng B.-Y., Chen H.-P., Wang S.-S. Improved Ant Colony Optimization Method for Single Batch-Processing Machine with Non-Identical Job Sizes. *Journal of System Simulation*. 2009;21(9):2687–2695.

20. Ramasubramanian M., Mathirajan M. A Mathematical Model for Scheduling a Batch Processing Machine with Multiple Incompatible Job Families, Non-identical Job dimensions, Nonidentical Job sizes, Non-agreeable release times and due dates. *Proceedings of the International Conference on Manufacturing, Optimization, Industrial and Material Engineering, MOIME 2013, 9–10 March 2013, Grand Hyatt, Bandung, Indonesia. IOP Conference Series: Materials Science and Engineering*. 2013. vol.46. DOI:10.1088/1757-899X/46/1/012013

21. Ogun B., Çigdem A.-U. Mathematical models for a batch scheduling problem to minimize earliness and tardiness. *Journal of Industrial Engineering and Management*. 2018;11(3):390–405. DOI:10.3926/jiem.2541

22. Kreipl S., Pinedo M. Planning and Scheduling in Supply Chains: An Overview of Issues in Practice. *Production and Operations Management*. 2004;13(1):77–92. DOI:10.1111/j.1937-5956.2004.tb00146.x

23. Krotov .V. Validation of a Multi-Level Programming Model for Scheduling Data Series Group Processing in a Pipelined Computing System with Constraints. *Proceedings of Irkutsk State Technical University.* 2016;1(108):35-47. (in Russ.)

24. Krotov K.V. Application of Genetic Algorithms for the Design of Effective and Comprehensive Schedules for the Processing of Data Packets in a Pipeline System at Specific Time Intervals. *Bulletin of the Voronezh State University. Series: System analysis and information technology*. 2020;1:71–86. (in Russ.) DOI:10.17308/sait.2020.1/2594

Сведения об авторе:

КРОТОВ Кирилл Викторович

кандидат технических наук, доцент, доцент кафедры «Информационные системы» Севастопольского государственного университета,
krotov_k1@mail.ru
https://orcid.org/0000-0002-9670-6141

DOI:10.31854/1813-324X-2020-6-3

УДК 003.26 DOI:10.31854/1813-324X-2020-6-3-92-99

Метод ЦВЗ для селективной аутентификации изображений, устойчивый к JPEG-сжатию, изменениям яркости и контрастности

А.Г. Жувикин[1]

[1]Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича, Санкт-Петербург, 193232, Российская Федерация
*Адрес для переписки: mail@zhuvikin.com

Информация о статье
Поступила в редакцию 19.06.2020
Принята к публикации 21.07.2020

Ссылка для цитирования: Жувикин А.Г. Метод ЦВЗ для селективной аутентификации изображений, устойчивый к JPEG-сжатию, изменениям яркости и контрастности // Труды учебных заведений связи. 2020. Т. 6. № 3. С. 92–99. DOI:10.31854/1813-324X-2020-6-3-92-99

Аннотация: *Одними из наиболее перспективных приложений для цифровых водяных знаков (ЦВЗ) являются системы селективной аутентификации изображений (САИ). Для реализации таких систем требуются методы вложения данных в покрывающее изображение со значительной емкостью. При этом метод вложения должен быть устойчив к классу естественных преобразованиям данной САИ-системы. В этой работе предлагается новый метод погружения ЦВЗ в изображение, который обладает большой емкостью вложения и является одновременно устойчивым к JPEG-сжатию, изменениям яркости и контрастности. Добиться такого результата удалось благодаря развитию метода вложения с использованием дискретного вейвлет-преобразования. Алгоритмом предусмотрено двухэтапное вложение, а также применение методик выравнивания и восстановления гистограммы. Экспериментальные результаты показали значительную устойчивость ЦВЗ к JPEG-сжатию, изменениям яркости и контрастности при высоком визуальном качестве изображения сразу после погружения по метрике PSNR.*

Ключевые слова: *яркость, контрастность, JPEG, дискретное вейвлет-преобразование, выравнивание гистограммы, системы селективной аутентификации изображений.*

Введение

Объем мультимедийного трафика в глобальной сети Интернет постоянно растет [1]. Это связано с тем, что человеческий мозг намного лучше адаптирован для восприятия такого рода информации. Несмотря на то, что передача мультимедийных данных требует большего объема трафика, чем обычные текстовые сообщения, люди все больше отдают им предпочтение для использования в социальных сетях. Важную роль занимают неподвижные цифровые изображения, используемые повсеместно: СМИ, государственные и частные Интернет-ресурсы, документооборот, фотохостинги и др. При этом неминуемо возникает риск искажения целого или части содержания таких изображений и, как следствие, введение потребителей изображений в преднамеренное заблуждение с различными целями. Подобные информационные угрозы, как правило, заключаются в подделке изображений при помощи графических редакторов, которые сейчас получили широкое распространение за счет дешевизны и простоты их использования.

Для предотвращения этих и других угроз были разработаны различные системы аутентификации изображений [2]. Системы точной аутентификации изображений (ТАИ) являются наиболее простыми и используют методы защиты данных при помощи криптографических средств. Несмотря на надежность ТАИ-системы имеют ряд ограничений для использования на повседневной основе благодаря возможности выполнения различных естественных преобразований над защищенными изображениями. В частности, для оптимизации дискового пространства серверов социальных сетей все изображения подлежат JPEG-сжатию с потерями, которые приводят к появлению небольших искажений, называемых артефактами сжатия. Хотя эти искажения и не меняют содержание изображений, но нарушают целостность электронной цифровой подписи (ЭЦП) в ТАИ-системах благодаря лавинообразному изменению значений криптографичес-

ких хеш-функций после изменения даже одного бита данных.

В качестве преодоления недостатка ТАИ-систем были разработаны различные алгоритмы компьютерного зрения, позволяющие оценить целостность содержания изображения для выполнения процедуры проверки подлинности. К таким методам относятся системы селективной аутентификации изображений (САИ) [2] и, в отличие от ТАИ-систем, являются устойчивыми к заранее выбранному классу естественных преобразований. Среди которых, кроме JPEG-сжатия, также выделяют изменение яркости и контрастности. Последние могут использоваться для улучшения визуального восприятия содержания изображения без нарушения его целостности. Популярным примером служит социальная сеть Instagram, где пользователь может корректировать изображение перед его отправкой, репостом и т. д.

САИ-системы часто используют цифровые водяные знаки (ЦВЗ) в качестве способа передачи данных необходимых для аутентификации [3–4]. ЦВЗ обладают существенными преимуществами по сравнению с передачей данных отдельным сообщением, например: предотвращают риск их нежелательной утери, позволяют скрыть сам факт наличия дополнительных данных от непосредственного потребителя, а также высвобождают дополнительный объем памяти, требуемый для их хранения и передачи.

Одним из главных препятствий в разработке САИ-систем при помощи ЦВЗ является выбор подходящего метода вложения. С одной стороны, он не должен приводить к значительному снижению визуального качества покрывающего изображения. С другой стороны, он должен быть устойчив к заранее выбранному классу естественных преобразований и обладать большой емкостью вложения, достаточной для выполнения процедуры селективной аутентификации. Более того, перечисленные выше требования являются взаимно противоречивыми.

В литературе описано достаточно много методов погружения для ЦВЗ, устойчивых к различным естественным преобразованиям. Однако, описанные методы обладают либо большой емкостью, но узконаправленной устойчивостью только к одному классу искажений, либо являются устойчивыми ко многим естественным искажениям, но не позволяют вложить достаточный объем данных. Так, устойчивость к изменению яркости и контрастности в работе [5] достигается за счет метода погружения в амплитуду и направление градиентов на границах объектов изображения. Авторы в работе [6] решают эту задачу при помощи вложения в коэффициенты дискретного косинусного преобразования (ДКП). Другие алгоритмы фокусируются на устойчивости ЦВЗ к JPEG-сжатию. Например, в [7] используется индексная модуляция коэффициентов ДКП, а в [8] предлагается симулировать возможные блочные искажения и вкладывать только в те области, где они наименьшие.

Искажения, вносимые естественными преобразованиями изображений

В отличие от преднамеренных искажений содержания изображений естественные преобразования являются следствием операций по уменьшению объема памяти, требуемой для хранения или улучшения визуального восприятия изображений. Предлагаемый метод вложения ЦВЗ предназначен для использования в САИ-системах, обладающих устойчивостью к следующим естественным преобразованиям.

JPEG-сжатие с потерями является классическим алгоритмом сжатия изображений за счет выполнения операции квантования коэффициентов ДКП областей 8×8 и последующим сжатием квантованных коэффициентов при помощи энтропийного кодирования. Искажения, возникающие при декодировании JPEG-изображений, называются артефактами сжатия.

Пусть $X = (x_{m,n}) \in \mathbb{R}^{N\times N}$ – матрица значений яркости отсчетов квадратного изображения размерами $N\times N$. Все изображение разбивается на непересекающиеся квадратные области 8×8, каждая из которых обрабатывается независимо. Операция JPEG-сжатия и восстановления изображения вместе с искажениями описывается в виде [9]:

$$\acute{X}_{k,l} = \text{ОДКП}\left(Q\left[\frac{\text{ДКП}(X_{k,l})}{Q}\right]\right), \qquad (1)$$

где $\acute{X}_{k,l}$ – матрица значений яркости отсчетов области $X_{k,l}$ размерами 8×8 после JPEG-сжатия и ее восстановления; [.] – операция округления до ближайшего целого; ОДКП – обратное ДКП, а $Q = (q_{m,n}) \in \mathbb{R}^{N\times N}$ – матрица квантования ДКП-коэффициентов, которая выбирается для каждого из возможных значений параметра качества JPEG-сжатия q и зависит от конкретной реализации JPEG-кодера.

Под *изменением яркости* изображения понимается следующая линейная операция:

$$\acute{X} = X(1 + b), \qquad (2)$$

где $\acute{X}$ – матрица значений яркости отсчетов после изменения яркости изображения; $b \in \mathbb{R}$ – параметр изменения яркости.

В реальных изображениях существуют ограничения на минимальное и максимальное возможные значения яркости одного отсчета. Для 8-битного цифрового изображения они равны $x_{\text{мин}} = 0$ и $x_{\text{макс}} = 2^8 - 1 = 255$, соответственно. Поэтому, несмотря на то, что операция изменения яркости (2) является линейной операцией, после ее применения возможно обрезание полученной гистограммы слева

или справа. На рисунке 1a заштрихованными областями показаны участки гистограммы изображения после применения операции изменения яркости с параметром b, выходящие за пределы динамического диапазона изображения.

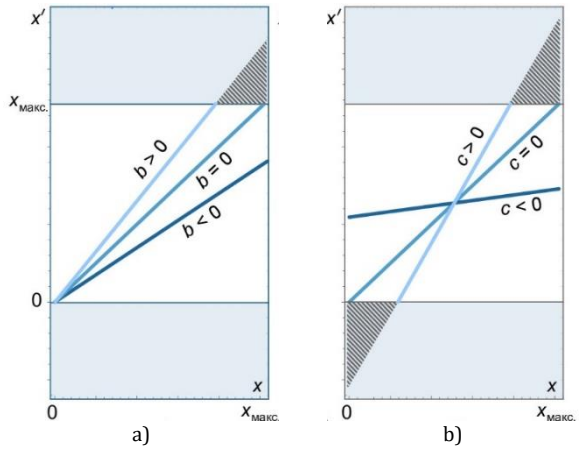


Рис. 1. Результат применения операций изменения яркости (a) и контрастности (b) на выходные значения отсчетов *x'* по входным значениям *x*. Заштрихованными областями показаны участки обрезания гистограмм

Fig. 1. An Effect of Brightness (a) and Contrast (b) Adjustments on Output Pixel Values X' Given by Source X. The Highlighted Areas Are Cut after Operations Performing

Цель *изменения контрастности*, в отличие от изменения яркости, заключается в том, чтобы сделать светлые участки изображения более или менее различимыми от темных. Поэтому такая операция может имеет разнонаправленный эффект на значения яркости отсчетов и выражается следующим образом:

$$\acute{X} = (1 + c)\left(X - \frac{x_{\text{макс}}}{2}\right) + \frac{x_{\text{макс}}}{2}, \qquad (3)$$

где $\acute{X}$ – матрица значений яркости отсчетов после изменения контрастности изображения; $c \in \mathbb{R}$ – параметр изменения контрастности. На рисунке 1b заштрихованными областями показаны участки гистограммы изображения после применения операции изменения контрастности с параметром c, выходящие за пределы динамического диапазона изображения.

Предложенный метод

В случае, если произошло обрезание гистограммы слева или справа, то часть данных ЦВЗ, вложенная в эти участки, может быть безвозвратно потеряна. Поэтому, строго говоря, невозможно создать метод вложения, который был бы устойчив к операциям изменения яркости или контрастности для любых значений параметров b и c, только для некоторого небольшого диапазона. Данное ограничение существует и для JPEG-сжатия. Успешное извлечение возможно только при небольших искажениях, которые возникают при некоторых допустимых пределах параметра качества сжатия q.

Предложенный метод погружения состоит из двух этапов. На первом этапе используется обычный метод вложения с использованием дискретного вейвлет-преобразования (ДВП). Далее используются обозначения областей коэффициентов ДВП в соответствии с рисунком 2.

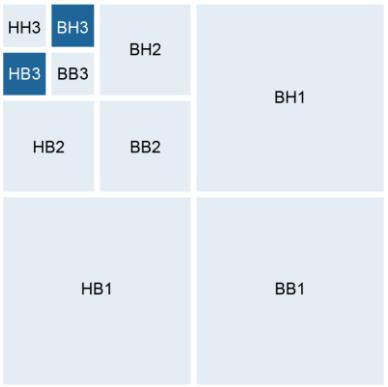


Рис. 2. Обозначение областей коэффициентов дискретного вейвлет-преобразования (ДВП)

Fig. 2. Denotation of the Discrete Wavelet Transform (DWT) Coefficients Areas

Для увеличения достоверности извлечения ЦВЗ данные предварительно кодируются с помощью кода коррекции ошибок, например, используя LDPC [10]. После чего каждый бит m_i данных ЦВЗ вкладывается в коэффициент s_i области HB3 (нижне- верхней) 3-уровневого ДВП по следующему правилу:

$$\acute{s}_i = \begin{cases} \gamma\left(\left[\frac{s_i}{\gamma}\right] + \frac{1}{4}\right), & m_i = 1 \\ \gamma\left(\left[\frac{s_i}{\gamma}\right] - \frac{1}{4}\right), & m_i = 0 \end{cases}, \qquad (4)$$

где $\acute{s}_i$ – коэффициент ДВП после вложения бита m_i; γ – глубина погружения.

Было показано, что данная техника позволяет добиться значительной устойчивости ЦВЗ к JPEG-сжатию [4]. Для включения изменения яркости и контрастности в класс допустимых естественных преобразований произведем модификацию алгоритма погружения путем добавления в него второго этапа.

На втором этапе изображение с погруженным ЦВЗ при помощи ДВП подвергается процедуре выравнивания гистограммы [11]. Это известная техника улучшения качества изображений, которая стремится привести гистограмму в соответствие с равномерным распределением. Отметим, что данная процедура является обратимым отображением значений яркости отсчетов исходного изображения на новые значения соответствующих отсчетов полученного изображения с «плоской» гистограммой. Пример результата применения операции выравнивания гистограммы для тестового изображения «Парусная лодка» приведен на рисунке 3.

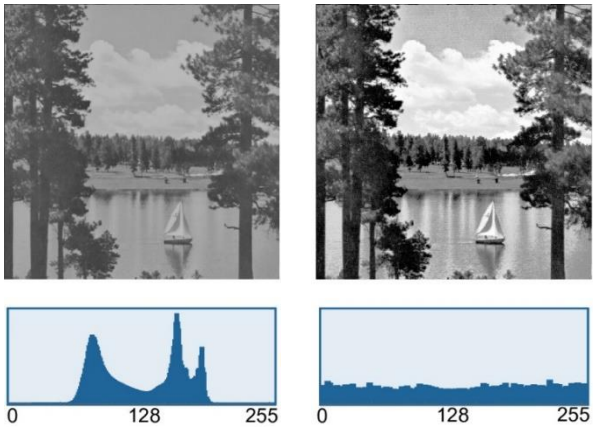


Рис. 3. Изображение «Парусная лодка» и гистограмма до (a) и после (b) применения операции выравнивания гистограммы изображения

Fig. 3. Image «Sailboat» with its Histogram before (a) and after (b) Histogram Equalization Operation

После выравнивания гистограммы выполняется дополнительное погружение методом ДВП по правилу (4–5), при этом погружаются те же самые, использованные на первом этапе, данные ЦВЗ, но в область ВНЗ (верхне-нижней) 3-уровневого ДВП. Полученное изображение с ЦВЗ подвергается процедуре восстановления гистограммы к ее исходной форме до погружения. Общая схема предлагаемого алгоритма погружения ЦВЗ приведена на рисунке 4.

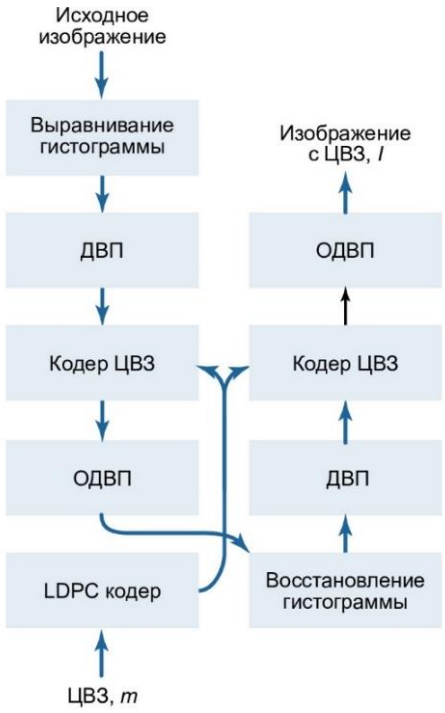


Рис. 4. Общая схема метода погружения ЦВЗ

Fig. 4. An Overall Scheme of the Watermark Embedding Procedure

Таким образом, за счет добавления второго этапа достигается устойчивость к небольшим изменениям яркости и контрастности.

Метод извлечения ЦВЗ выполняется в порядке обратному схеме погружения. Сначала принятое изображение с возможными искажениями копируется. После чего, одна копия раскладывается на области ДВП коэффициентов сразу, а вторая – после выполнения процедуры предварительного выравнивания гистограммы. Каждый коэффициент из областей НВЗ и ВНЗ 3-уровневого ДВП поступает на декодер ЦВЗ, который принимает решение по следующему алгоритму [4]:

$$\acute{m}_i = \begin{cases} 1, & \text{если } \acute{s}_i - \gamma\left[\frac{\acute{s}_i}{\gamma}\right] \geq 0, \\ 0, & \text{если } \acute{s}_i - \gamma\left[\frac{\acute{s}_i}{\gamma}\right] < 0, \end{cases}, \qquad (5)$$

где $\acute{m}_i$ – бит, извлеченный из коэффициента $\acute{s}_i$.

После извлечения из областей НВЗ и ВНЗ выполняется процедура декодирования LDPC [10]. Общая схема извлечения ЦВЗ из принятого изображения с возможными искажениями отображена на рисунке 5.

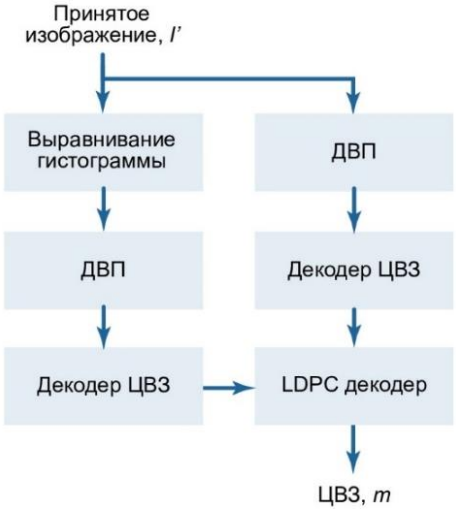


Рис. 5. Общая схема метода извлечения ЦВЗ из принятого изображения с возможными искажениями

Fig. 5. An Overall Scheme of the Watermark Data Extraction Procedure from Received Image with Possible Modifications

В качестве примера САИ-системы, для которой может быть использован предложенный алгоритм устойчивого погружения, можно привести метод из [4]. По исходному содержанию изображения вычисляется значение перцептивной хеширующей функции, которая обладает свойством изменяться незначительно при применении естественных преобразований и значительно при преднамеренных. В качестве перцептивной хеширующей функции может быть выбрана, например, последовательность центральных конечных разностей. Формирование данных ЦВЗ m включает в себя вычисление:

– вектора пертурбаций p, сформированного при помощи алгоритма 3-битного квантования [4] по значению перцептивной хеширующей функции содержания исходного изображения;

– ЭЦП S квантованного значения перцептивной хеширующей функции, вычисленной при помощи секретного ключа d владельца изображения.

Битовые последовательности p и S вместе погружаются в виде данных ЦВЗ при помощи предложенного устойчивого алгоритма. После передачи изображения и выполняется процедура селективной аутентификации. Извлеченный вектор p участвует в процедуре восстановления значения перцептивной хеширующей функции содержания исходного изображения по принятому изображению с возможными искажениями при помощи алгоритма 3-битного квантования. После восстановления квантованного значения перцептивной хеширующей функции выполняется классическая процедура проверки ЭЦП при помощи открытого ключа k владельца изображения. В случае, если подписи совпадают, САИ-система признает изображение верным с точностью до возможного применения естественных преобразований, в противном случае – признается подделкой. При этом использование стойкой ЭЦП, вырабатываемой по секретному ключу владельца, позволяет защитить ЦВЗ от преднамеренных атак нарушителей при помощи, например, атаки «подмены» ЦВЗ.

Для подтверждения того, что предложенный алгоритм решает поставленные задачи, был произведен ряд экспериментов, в которых была использована база SIPI из 38 цифровых изображений с различным содержанием [12]. В результате экспериментов было установлен диапазон оптимальных значений параметра $\gamma \in [0{,}05; 0{,}15]$, при котором достигается высокое визуальное качество изображения после вложения ЦВЗ и, одновременно, приемлемые показатели устойчивости к выбранному классу естественных преобразований: JPEG-сжатию, изменению яркости и контрастности.

Устойчивость ЦВЗ к изменениям яркости, контрастности и JPEG-сжатию была протестирована на каждом из изображений. На рисунках 6–7 отображены зависимости BER от параметров изменения яркости b и контрастности c при $\gamma = 0{,}11$. Видно, что в целом устойчивость к уменьшению яркости и контрастности выше, чем к их увеличению. Это связано с тем, что при отрицательных значениях b и c не происходит отрезание частей гистограммы слева или справа. Метод обеспечивает BER ≤ 5 % при $b \in [-0{,}3; 0{,}2]$ или $c \in [-0{,}3; 0{,}15]$. На рисунке 8 приведены результаты тестирования устойчивости предложенного метода ЦВЗ к JPEG-сжатию при оптимальном значении параметра $\gamma = 0{,}11$. По результатам эксперимента видно, что алгоритм обеспечивает BER ≤ 5 % для $q \geq 70$.

Отметим, что BER ≤ 5 % надежно исправляются при помощи кода коррекции ошибок LDPC со скоростью ½. Таким образом, общая полезная емкость C вложения для квадратного изображения размерами N^2 может быть рассчитана по выражению:

$$C = \left\lfloor \frac{1}{2}\left(\frac{N}{2^3}\right)^2 \right\rfloor \text{ бит}, \qquad (6)$$

где $\lfloor x \rfloor$ – округление числа x до наибольшего целого меньшего x (например, для изображений размерами 512×512 общая полезная емкость вложения с поправкой на избыточность для кода LDPC со скоростью ½ составляет 2048 бит, что является достаточным для реализации САИ-системы на его основе). Например, передача ЭЦП на основе алгоритма ECDSA с эллиптической кривой (secp256k1) потребует 512 бит [13]. Остальные 1536 бит могут быть выделены для реализации алгоритма 3-битного квантования перцептивной хеш-функции на основе 1536/3 = 512 коэффициентов ДВП.

Другой важной характеристикой любого метода ЦВЗ является метрика качества изображения сразу после погружения данных. Для оценки предложенного алгоритма использовалась распространенная метрика PSNR (*от англ.* Peak Signal-to-Noise Ratio), вычисляемая по выражению [14]:

$$\text{PSNR} = \log_{10}\left[\frac{255^2}{\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}(I[i,\ j] - \acute{I}[i,\ j])^2}\right] \text{ дБ}, \qquad (7)$$

где $\acute{I}$ – изображение сразу после погружения ЦВЗ.

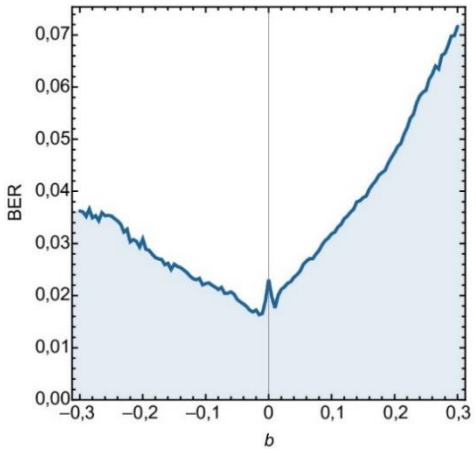


Рис. 6. Зависимость BER от параметра изменения яркости b при $\gamma = 0{,}11$

Fig. 6. The Dependence of BER on Brightness Adjustment Parameter b with γ = 0,11

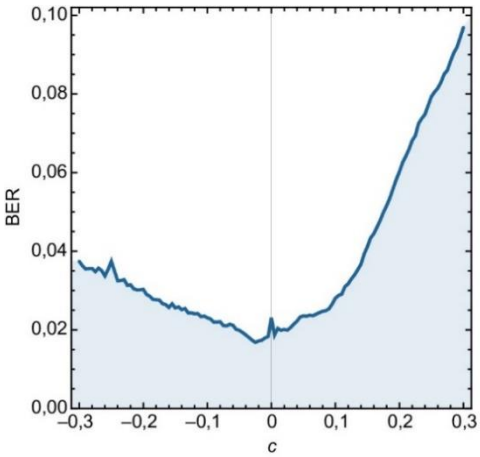


Рис. 7. Зависимость BER от параметра изменения контрастности c при $\gamma = 0{,}11$

Fig. 7. The Dependence of BER on Contrast Adjustment Parameter c with γ = 0,11

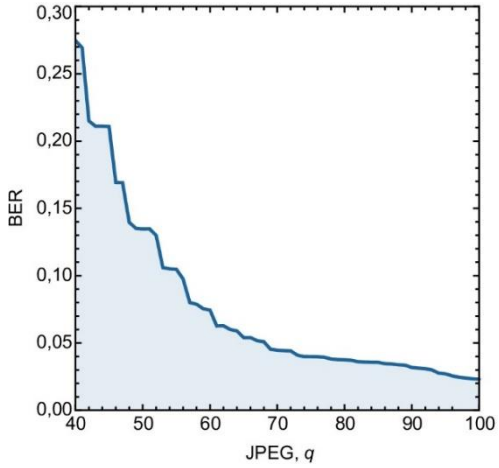


Рис. 8. Зависимость BER от параметра JPEG-сжатия *q* при γ = 0,11

Fig. 8. The Dependence of BER on JPEG Compression Parameter q with γ = 0,11

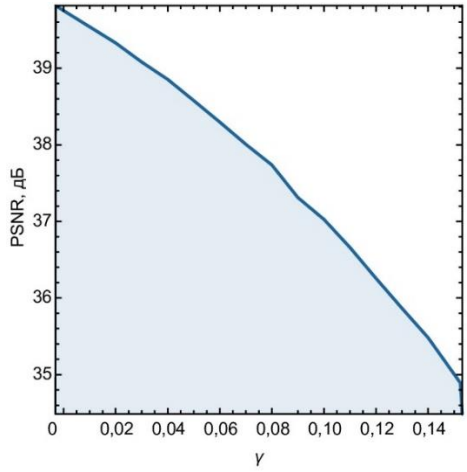


Рис. 9. Зависимость значений метрики PSNR от глубины погружения γ

Fig. 9. The Dependence of visual Quality Metric PSNR on Embedding Parameter γ

На рисунке 9 отображена зависимость значений метрики PSNR от глубины погружения γ, а на рисунке 10 приведен пример изображения с ЦВЗ при γ = 0,11. Как видно, предложенный метод показывает высокое визуальное качество после вложения.

Рис. 10. Тестовое изображение «Парусная лодка» с ЦВЗ

Fig. 10. A Watermarked Version of Test Image «Sailboat»

Заключение

Предложенный новый метод вложения ЦВЗ в цифровые изображения позволил получить устойчивость одновременно к сжатию JPEG, изменению яркости и контрастности. Достичь этого позволило использование двухэтапного вложения с использованием ДВП, процедур выравнивания и восстановления гистограммы.

Несмотря на то, что извлечение бит ЦВЗ выполняется с некоторой вероятностью ошибки BER, исходные данные могут быть полностью восстановлены благодаря предварительному использованию помехоустойчивого кодирования. Например, LPDC, для широких значений параметров сжатия *q*, изменения яркости *b* и изменения контрастности *c*. Общая емкость вложения, рассчитываемая по (6), позволяет использовать предложенный метод ЦВЗ для САИ-систем на основе устойчивых перцептивных хеширующих функций, стойкой криптографической подписи и 3-битного квантования.

Список используемых источников

1. Cisco Annual Internet Report (2018–2023). Available from: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.pdf (дата обращения 05.08.2020)

2. Haouzia A., Noumeir R. Methods for image authentication: a survey // Multimed Tools and Applications. 2008. Vol. 39. Iss. 1-46. DOI:10.1007/s11042-007-0154-3

3. Korzhik V., Zhuvikin A., Morales-Luna G. Selective image authentication tolerant to JPEG compression // Proceedings of the 6th International Conference on Information, Intelligence, Systems and Applications (IISA, Corfu, Greece, 6–8 July 2015). IEEE, 2015. DOI:10.1109/IISA.2015.7388076

4. Zhuvikin A., Korzhik V., Morales-Luna G. Semi-Fragile Image Authentication based on CFD and 3-Bit Quantization // Indian Journal of Science and Technology. 2016. Vol. 9. Iss. 48. PP. 1–7. DOI:10.17485/ijst/2016/v9i48/109167

5. Daofu G., Fenlin L., Xiangyang L., Ping W. Robust Image Watermarking against Brightness and Contrast Change // Proceedings of the 4th International Conference on Multimedia Information Networking and Security (Nanjing, China, 2–4 November 2012). IEEE, 2012. PP. 995–998. DOI:10.1109/MINES.2012.252

6. Li X.-D. Image digital watermarking algorithm in DCT domain for resisting brightness-and-contrast adjusting attack // Journal of Optoelectronics. Laser. 2013. Vol. 6. URl: http://en.cnki.com.cn/Article_en/CJFDTotal-GDZJ201306027.htm (дата обращения 03.08.2020)

7. Muñoz-Ramirez D.-O., Ponomaryov V., Reyes-Reyes R., Kyrychenko V., Pechenin O., Totsky A. A robust watermarking scheme to JPEG compression for embedding a color watermark into digital images // Proceedings of the 9th International Conference on Dependable Systems, Services and Technologies (DESSERT, Kiev, Ukraine, 24–27 May 2018). IEEE, 2018. PP. 619–624. DOI:10.1109/DESSERT.2018.8409206
8. Nguyen P., Beghdadi A., Luong M. Perceptual watermarking robust to JPEG compression attack // Proceedings of the 5th International Symposium on Communications, Control and Signal Processing, Rome, Italy, 2–4 May 2012). IEEE, 2012. PP. 1–4. DOI:10.1109/ISCCSP.2012.6217798
9. Wang X., Feng X., He N., Jia Q., Zou Y. JPEG image compression algorithm analysis and local optimization // Proceedings of the International Conference on Advanced Management Science (ICAMS 2010, Chengdu, China, 9–11 July 2010). IEEE, 2010. PP. 5–7. DOI:10.1109/ICAMS.2010.5553016
10. Gallager R.G. Low-density parity-check codes. Cambridge: M.I.T. Press, 1963.
11. Zhu Y., Huang C. Histogram equalization algorithm for variable gray level mapping // Proceedings of the 8th World Congress on Intelligent Control and Automation (Jinan, China, 7–9 July 2010). IEEE, 2010. PP. 6022–6025. DOI:10.1109/WCICA.2010.5554587
12. The USC-SIPI Image Database // University of Southern California. Viterbi School of Engineering. URL: http://sipi.usc.edu/database (дата обращения 17.06.2020)
13. Brown D. The Exact Security of ECDSA // In Advances in Elliptic Curve Cryptography. Cambridge: Cambridge University Press, 2000.
14. Huynh-Thu Q., Ghanbari M. Scope of validity of PSNR in image/video quality assessment // Electronics letters. 2008. Vol. 44. Iss. 13. PP. 800–801. DOI:10.1049/el:20080522

* * *

A Watermarking Method for Selective Image Authentication Tolerant to Jpeg Compression, Brightness and Contrast Adjustments

A. Zhuvikin[1]

[1]The Bonch-Bruevich Saint-Petersburg State University of Telecommunications,
St. Petersburg, 193232, Russian Federation

Article info
DOI:10.31854/1813-324X-2020-6-3-92-99
Received 19th June 2020
Accepted 21th July 2020

For citation: Zhuvikin A. A Watermarking Method for Selective Image Authentication Tollerant to JPEG Compression, Brightness and Contrast Adjustments. *Proc. of Telecom. Universities.* 2020;6(3):92–99. (in Russ.) DOI:10.31854/1813-324X-2020-6-3-92-99

Abstract: *One of the most promising application of the digital watermarking is the selective image authentication (SIA) systems. In order to implement such a system one requires an embedding algorithm with an appropriate capacity. In addition, an embedding method is to be robust for the class of non-malicious manipulations which the SIA system is designed for. We propose the new method which has a significant embedding capacity while still being tolerant to JPEG compression, brightness and contrast adjustments. This was possible due to the extension of the well-known discrete wavelet transform embedding technique. We propose two-step embedding scheme and the use of image histogram equalisation and recovering operations. The experiment results show acceptable tolerance to JPEG compression, brightness and contrast adjustments with good visual quality in terms of PSNR just after embedding.*

Keywords: *brightness, contrast, JPEG, discrete wavelet transform, histogram equalisation, selective image authentication systems.*

References

1. *Cisco Annual Internet Report (2018–2023)*. Available from: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.pdf [Accessed 5th August 2020]
2. Haouzia A., Noumeir R. Methods for image authentication: a survey. *Multimed Tools and Applications*. 2008;39(1-46). DOI:10.1007/s11042-007-0154-3
3. Korzhik V., Zhuvikin A., Morales-Luna G. Selective image authentication tolerant to JPEG compression. *Proceedings of the 6th International Conference on Information, Intelligence, Systems and Applications, IISA, 6–8 July 2015, Corfu, Greece*. IEEE; 2015. DOI:10.1109/IISA.2015.7388076
4. Zhuvikin A., Korzhik V., Morales-Luna G. Semi-Fragile Image Authentication based on CFD and 3-Bit Quantization. *Indian Journal of Science and Technology*. 2016;9(48):1–7. DOI:10.17485/ijst/2016/v9i48/109167
5. Daofu G., Fenlin L., Xiangyang L., Ping W. Robust Image Watermarking against Brightness and Contrast Change. *Proceedings of the 4th International Conference on Multimedia Information Networking and Security, 2–4 November 2012, Nanjing, China*. IEEE; 2012. p.995–998. DOI:10.1109/MINES.2012.252
6. Li X.-D. Image digital watermarking algorithm in DCT domain for resisting brightness-and-contrast adjusting attack. *Journal of Optoelectronics. Laser*. 2013;6. Available from: http://en.cnki.com.cn/Article_en/CJFDTotal-GDZJ201306027.htm [Accessed 3rd August 2020]
7. Muñoz-Ramirez D.-O., Ponomaryov V., Reyes-Reyes R., Kyrychenko V., Pechenin O., Totsky A. A robust watermarking scheme to JPEG compression for embedding a color watermark into digital images. *Proceedings of the 9th International Conference on Dependable Systems, Services and Technologies, DESSERT, 24–27 May 2018, Kiev, Ukraine*. IEEE; 2018. p.619–624. DOI:10.1109/DESSERT.2018.8409206
8. Nguyen P., Beghdadi A., Luong M. Perceptual watermarking robust to JPEG compression attack. *Proceedings of the 5th International Symposium on Communications, Control and Signal Processing, 2–4 May, Rome, Italy*. IEEE; 2012. p.1–4. DOI:10.1109/ISCCSP.2012.6217798
9. Wang X., Feng X., He N., Jia Q., Zou Y. JPEG image compression algorithm analysis and local optimization. *Proceedings of the International Conference on Advanced Management Science, ICAMS 2010, 9–11 July 2010, Chengdu, China*. IEEE; 2010. p.5–7. DOI:10.1109/ICAMS.2010.5553016
10. Gallager R.G. *Low-density parity-check codes*. Cambridge: M.I.T. Press; 1963.
11. Zhu Y., Huang C. Histogram equalization algorithm for variable gray level mapping. *Proceedings of the 8th World Congress on Intelligent Control and Automation, 7–9 July 2010, Jinan, China*. IEEE; 2010. p.6022–6025. DOI:10.1109/WCICA.2010.5554587
12. University of Southern California. Viterbi School of Engineering. *The USC-SIPI Image Database*. Available from: http://sipi.usc.edu/database [Accessed 17th June 2020]
13. Brown D. The Exact Security of ECDSA. *In Advances in Elliptic Curve Cryptography*. Cambridge: Cambridge University Press; 2000.
14. Huynh-Thu Q., Ghanbari M. Scope of validity of PSNR in image/video quality assessment. *Electronics letters*. 2008;44(13):800–801. DOI:10.1049/el:20080522

Сведения об авторе:

ЖУВИКИН Алексей Георгиевич

аспирант кафедры защищенных систем связи Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича, mail@zhuvikin.com

https://orcid.org/0000-0002-5120-3006

УДК 539.12 DOI:10.31854/1813-324X-2020-6-3-100-106

Прогнозирование ресурсов облачных сервисов на основе мониторинговой системы с открытым кодом

К.Н. Кучерова[1*]

[1]Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, 195251, Российская Федерация
[*]Адрес для переписки: kristina.mylife@gmail.com

Информация о статье
Поступила в редакцию 05.08.2020
Принята к публикации 22.09.2020

Ссылка для цитирования: Кучерова К.Н. Прогнозирование ресурсов облачных сервисов на основе мониторинговой системы с открытым кодом // Труды учебных заведений связи. 2020. Т. 6. № 3. С. 100–106. DOI:10.31854/1813-324X-2020-6-3-100-106

Аннотация: *В статье описан универсальный подход к мониторингу хранилищ данных глобально распределенных вычислительных комплексов, что позволяет автоматизировать создание новых метрик в системе и прогнозировать их поведение для конечного пользователя. Так как существующие мониторинговые программные продукты обеспечивают готовую схему для мониторинга только системных метрик, таких как использование оперативной памяти, процессора, внешних дисков и сетевого траффика, но не предлагают решения для бизнес-функций, то IT-компаниям приходится проектировать специализированные структуры баз данных. Предложенная в статье структура данных для хранения мониторинговой статистической информации универсальна и позволяет экономить ресурсы при организации мониторинга баз данных в масштабе глобально распределенных вычислительных комплексов. Целью работы является разработка универсальной модели для мониторинга и прогнозирования хранилищ данных глобально распределенных вычислительных комплексов и оценка ее адекватности реальным условиям эксплуатации.*

Ключевые слова: *облачная система, база данных, мониторинг, системная метрика, прогнозирование, корреляционная функция.*

Введение

Как известно, контролировать системные метрики нужно для того, чтобы быстро реагировать на технические проблемы в системе, например, на отсутствие свободного места на диске, или перегруженность центального процессора (также, центрального процессорного устройства – ЦПУ), и тем самым поддерживать высокий уровень доступности облачных сервисов для конечных пользователей. Однако далеко не все понимают значимость мониторинга бизнес-метрик, которые позволяют обнаружить проблему в бизнес-логике сервисов, в то время как с технической точки зрения встроенные системные метрики будут показывать, что с системой все в порядке.

Например, такие бизнес-метрики, как «Количество Интернет-соединений клиентов» и «Количество транзакций в минуту», могут выявить неожиданное сочетание большого количества пользователей и подозрительно маленького количества операций, что, скорее всего, будет указывать на неработоспособность сервисов у большей части клиентов. Эту проблему и подобные ей невозможно обнаружить с помощью стандартных системных метрик, а только с помощью специализированных бизнес-метрик.

Правильно подобранные бизнес-метрики позволяют обнаружить ошибки в логике работы веб-приложения или, наоборот, указать на удачные решения, когда потребности конечных пользователей растут после выпуска новой версии программного продукта.

В данной статье предложен универсальный подход к мониторингу баз данных (БД) в составе глобально распределенных вычислительных комплексов (ГРВК), который позволяет автоматизировать добавление новых бизнес-метрик без изменения структуры данных и написания новых процедур взаимодействия между БД и системой мониторинга.

Анализ существующих подходов и постановка задачи мониторинга ГРВК

Объектом исследования является международная телекоммуникационная компания Distillery, США [1], где в качестве БД используется MS SQL Server 2016 Enterprise Edition. Предметом исследования является система мониторинга с открытым исходным кодом Zabbix 3.0 [2], позволяющая вносить изменения в программный продукт без ущерба последующей миграции на новые версии. Но предложенный универсальный подход может быть использован для любой другой системы мониторинга и БД реляционного типа.

Существующий подход к созданию новых метрик в Zabbix и других аналогичных мониторинговых системах имеет крупный недостаток. Он заключается в отсутствии автоматизации и необходимости ручного создания специальных таблиц БД для хранения описания метрик и разработке SQL-запросов либо хранимых процедур для расчета их значений. При таком подходе каждое добавление новой метрики связано с модификацией структуры БД и написанием специализированных SQL-запросов, а также интеграцией с существующей системой мониторинга.

Заказчиком и потребителем системных метрик обычно являются только технические отделы компании, и их мониторинг используется для проверки доступности и работоспособности всей системы в целом в режиме 24/7 [3]. В случае с мониторингом бизнес-метрик ситуация другая, т. к. их наблюдают и анализируют другие специалисты компании. Это накладывает дополнительные требования к задаче, в том числе необходимость более гибко и быстро добавлять в систему новые бизнес-метрики по требованию отдельных групп заказчиков. Например, аналитический отдел может попросить в дополнение к существующей метрике количества транзакций в день добавить новое распределение по времени суток, отдельным заказчикам, регионам, типам операций и другим параметрам.

Системный мониторинг является стандартизированным, т. к. все системы используют серверы похожей архитектуры, где есть одинаковые компоненты: дисковое пространство, оперативная память, быстродействие и количество ядер ЦПУ, в том числе распространенные метрики, которые можно получать из операционной системы (ОС). По этой причине системный мониторинг поставляется в стандартном пакете, причем с использованием встроенных шаблонов ОС в случае Zabbix, и достаточно просто установить систему мониторинга и получить базовые системные метрики автоматически без дополнительной настройки [4]. С бизнес-метриками все иначе, поскольку все бизнес-системы отличаются друг от друга по архитектуре и реализации, нет никаких общих стандартов для получения метрик внешними системами. Кроме того, методы расчета бизнес-метрик могут отличаться от сервиса к сервису, поэтому на данном этапе развития IT-индустрии и стандартов шаблонный мониторинг бизнес-метрик не кажется реализуемым.

Идеальным подходом для бизнес-мониторинга была бы возможность добавлять новые метрики без необходимости внесения изменений в саму систему, а именно через настройку мониторинга без модификации структуры БД. К сожалению, в существующих мониторинговых решениях такая возможность отсутствует.

Интересным примером гибкой реализации подхода к бизнес-мониторингу является система AppDynamics [5], которая позволяет добавлять нужные метрики по бизнес-транзакциям. Т. к. AppDynamics осуществляет мониторинг веб-приложений, то в системе можно настроить сбор метрик по событию и параметрам приложения и таким образом создавать необходимые срезы для просмотра состояния бизнес-транзакций без внесения изменений в код приложения, а также без необходимости дорабатывать интеграцию с системой мониторинга. Тем не менее, продукт AppDynamics не является полностью универсальным решением для всех типов веб-приложений, а затраты на его использование трудно согласовать между заказчиком и исполнителем.

Многие IT-компании признают необходимость отслеживания бизнес-метрик и описывают их преимущества, но сами не предоставляют готовую схему реализации. Так, система PagerDuty [6] является ведущим решением по оповещению и отслеживанию инцидентов в реальном режиме времени 24/7 с привлечением мобильных устройств обслуживающего персонала. Но предоставление информационных услуг ограничивается веб-интерфейсом и его связью с любыми типами компьютеров клиентов, а бизнес и прочие метрики наблюдения и оценки работоспособности систем остаются на совести конечных пользователей.

Популярное решение для мониторинга бизнес-метрик завоевывает Prometheus [7], т. к. имеет высокую производительность и масштабируемость, а также готовые шаблоны для мониторинга системных метрик. Норвежская компания FINN рассказывает о своем опыте применения Prometheus для мониторинга бизнес-метрик [8], из которого видно, что несмотря на удобства настройки метрик, сам бизнес-мониторинг требует реализации и адаптации для каждой компании.

На российском рынке телекоммуникаций и технологий Internet of Things (IoT) лидером мониторинга бизнес-метрик является разработка компании Saymon [9]. Ее преимущество состоит в возможности конфигурации объекта и выборки данных из других систем мониторинга, создании

экранов с консолидированной информацией и настройки связей между объектами, что позволяет получить более полную картину изменения статусной модели объекта в системе с течением времени.

Таким образом, ни одна из существующих мониторинговых систем не является универсальной с точки зрения контроля бизнес-метрик, которые в разных IT-компаниях имеют свою бизнес-логику сбора и анализа данных. Для дальнейшего решения поставленной задачи необходимо разработать информационную модель в виде универсальной структуры данных БД для хранения любых типов бизнес-метрик, провести анализ ее адекватности условиям эксплуатации ГРВК и определить область применения для мониторинга IT-объектов.

Описание мониторинга системных метрик

Универсальная система мониторинга системных и бизнес-метрик разработана и внедрена в международной компании Distillery, США [1], которая входит в аналитический отчет Gartner [10] в своем секторе по синхронизации файловых ресурсов (рисунок 1), имеет несколько миллиардов файлов на обслуживании и корпоративных клиентов с численностью более 300 тыс. человек в США и Европе.

Рис. 1. Квадрант Гартнер по корпоративным системам обмена и синхронизации файлов (июнь 2018)

Fig. 1. Gartner Magic Quadrant for Content Collaboration Platforms (as of June 2018)

Актуальность разработки вызвана возросшими требованиями корпоративных клиентов к показателям доступности и работоспособности сервисов. Жесткое соглашение Service Level Agreement (SLA) предполагает доступность сервисов для клиентов 99,99 % в режиме 24/7, в противном случае в результате каждого отказа в обслуживании компания вынуждена выплачивать своим контрагентам неустойки.

Важным условием новой разработки является ее интеграция с существующим мониторингом системных метрик Zabbix, который допускает изменение исходного кода и модификацию структуры БД. Для этого использованы технологии Microsoft, серверы C# .net, БД MS SQL, 14 активных серверов и 14 – в режиме зеркалирования данных. Постоянная нагрузка в пиковые часы составляет порядка 400 тыс. запросов в минуту.

На рисунке 2 приведен пример системной метрики мониторинга свободного места на дисках серверов БД [11], а на рисунке 3 показана логика соблюдения соотношения исторического периода и горизонта прогнозирования для своевременного реагирования на окончание свободного места на диске [12].

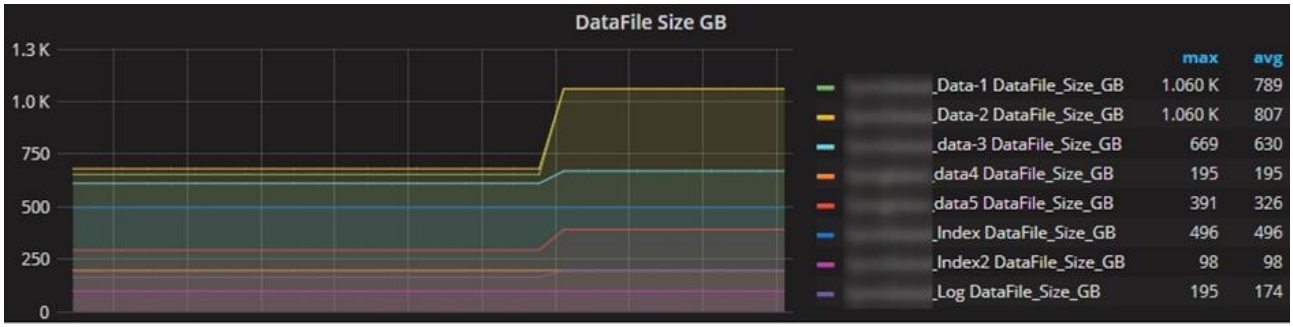
Разработка универсальной структуры БД мониторинга бизнес-метрик

Список бизнес-метрик ГРВК, которые необходимо хранить в БД и добавить в мониторинг в дополнение к стандартным системным метрикам, представлен в таблице 1.

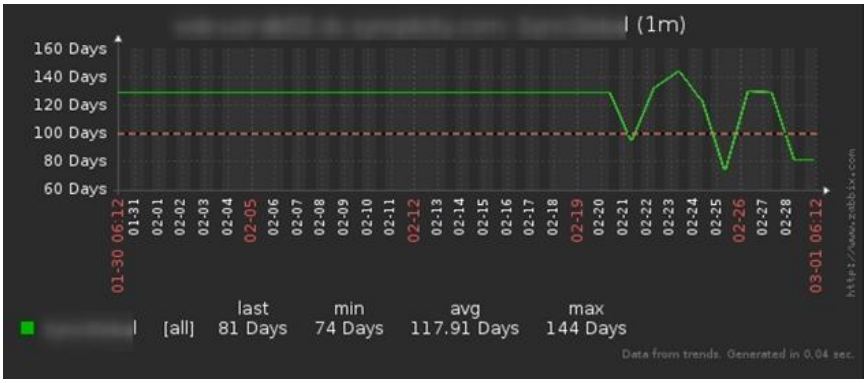
ТАБЛИЦА 1. Список бизнес-метрик мониторинга ГРВК

TABLE 1. The List of Business Metrics for Monitoring Globally Distributed Cloud Computing System

№ п/п	Название	Ед. изм.	Нижний порог срабатывания	Верхний порог срабатывания
1.	Количество загруженных файлов	%	1	20
2.	Количество удаленных файлов	%	5	30
3.	Всего файлов в системе	шт.	–	–
4.	Всего папок и файлов с распределенным доступом	шт.	–	–
5.	Количество компаний-клиентов	шт.	–	–
6.	Количество компаний-клиентов с несколькими пользователями	%	50	–
7.	Общее количество уникальных пользователей с корпоративной подпиской	%	40	–
8.	Общее количество папок в системе	шт.	–	–
9.	Количество пользователей в разрезе по серверам	шт.	–	–
10.	Количество папок по серверам	шт.	–	–



a)



b)

Рис. 2. График использования свободного места на дисках серверов (а) БД и их прогноз (b)

Fig. 2. Graphic of Free Disk Space on DB Servers (a) and Its Prediction (b)

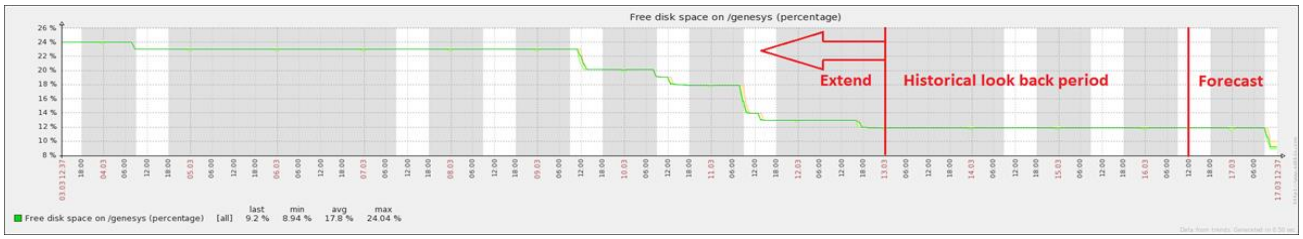


Рис. 3. Оценка соотношения исторического периода и горизонта прогнозирования

Fig. 3. Evaluation of Relationship between Historical Period and Forecast Horizon

Бизнес-метрики «Количество компаний-клиентов» и «Количество компаний-клиентов с несколькими пользователями» кажутся одинаковыми, но крайне важно собирать различные сведения по такому параметру, как количество пользователей. Если пользователь зарегистрировался, но потом не пользовался сервисом, тогда их фактическое количество не соответствует формальному, и нагрузка на систему будет меньше. Для реализации мониторинга бизнес-метрик таблицы 1 создана универсальная структура из четырех таблиц БД.

1) MetricDescription – таблица со списком метрик, которые необходимо собирать.

Структура таблицы:

MetricDescriptionId – идентификатор описания; MetricShortName – сокращенное название, используемое для обозначения метрики в процедурах сбора метрик;

MetricUserName – название, которое нужно отображать пользователю;

MetricDescription – описание бизнес-смысла метрики;

MetricGroupByColumn – поле группировки, но для метрик, состоящих из одного параметра, оно не задается;

MetricDBSource – БД-источник данных по метрике;

MetricSelectType – тип выборки либо выборка по расписанию, либо метрика, которая собирается в связи с каким-то событием в системе;

MetricOrder – порядок сортировки метрики.

2) MetricData – таблица для хранения данных по собранным метрикам.

Поля таблицы:

MetricDataId – идентификатор строки;

MetricServerId – идентификатор сервера, на котором была собрана метрика;

MetricGroupId – идентификатор группы, если метрика групповая;

MetricDescriptionId – ссылка на идентификатор описания метрики;

MetricGroupByColumn – значения идентификатора группирующего поля (заполнено, только если в MetricDescription задан MetricGroupByColumn);

MetricGroupByColumnDescription – название группирующего поля (заполнено, только если в MetricDescription задан MetricGroupByColumn);

MetricDataBigInt – значения метрики;

DateCreatedUTC – дата сбора метрики;

OrderData – порядок сортировки.

Также в структуре БД есть две таблицы для настройки отчета по значениям метрик:

3) MetricReport – заголовок отчета;

4) MetricReportContent – список метрик, которые входят в отчет.

Сбор метрик осуществляется по расписанию, которое задает администратор БД. В соответствии с этим расписанием вызывается процедура MetricsCalculateAll, в которой идут обращения к базам данных, указанным в MetricDBSource. Для каждой БД создана специальная процедура spMetricsCalculate, в которую передается список собираемых метрик. Внутри процедуры осуществляется подсчет только тех метрик, которые включены в список. Для определения наличия метрики в списке используется ShortName, так как это более читабельно, чем просто идентификатор метрики, что позволяет уменьшить количество ошибок.

Если метрика есть в списке, она рассчитывается заранее определенным способом, и результаты складываются во временную таблицу БД, которая возвращается в качестве результата в конце процедуры spMetricsCalculate, и таким образом результаты передаются в вызывающую процедуру MetricsCalculateAll. Когда результаты процедуры приходят в БД для хранения метрик, они записываются в таблицу MetricData.

Затем из мониторинговой системы Zabbix периодически вызывается выдающая данные по метрикам и их краткое описание в систему мониторинга процедура MetricsGetForMonitoring. Далее в системе мониторинга можно настроить необходимый вариант отображения метрик в виде графика или таблицы.

Практическая реализация системы мониторинга бизнес-метрик

На рисунке 4 приведен пример мониторинга бизнес-метрик «Количество загруженных файлов» в процентах от «Всего файлов в системе». Для удобства визуализации полученный процент умножается на 1 000 000. В результате на графике отмечено падение загрузки файлов в систему, в то время как все системные метрики были в норме. Как видно из графика, такое количество загруженных файлов не характерно для этого периода суток и свидетельствует о какой-то возможной аномалии в системе. Позднее было выявлено, что это произошло из-за ошибки в логике приложения.

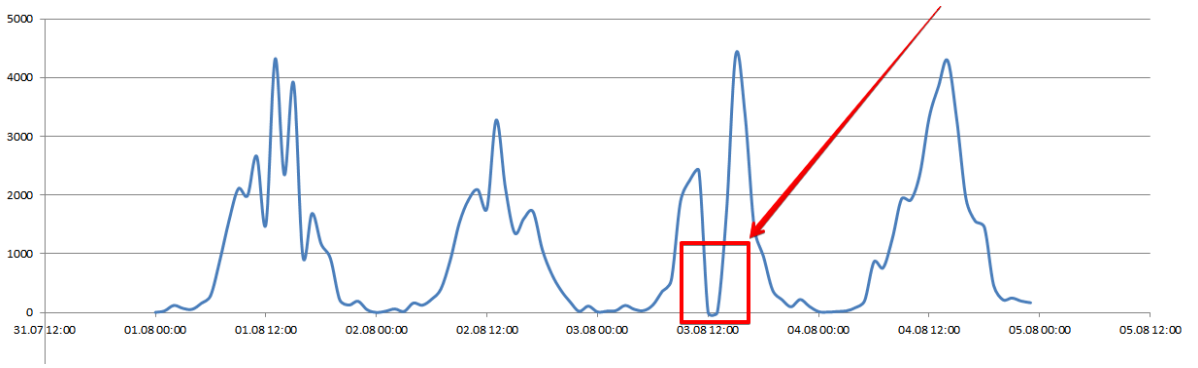


Рис. 4. Пример обнаружения аномалии в бизнес-метрике «Количество загруженных файлов, %»

Fig. 4. Example of an Anomaly Detection in the Business Metric "Uploaded Files, %"

Описанный подход к мониторингу бизнес-метрик реализован на практике в международной компании Distillery, США [1]. Это позволило добавить в мониторинг новые бизнес-метрики и упростило обнаружение и анализ причин появления аномальных ситуаций. В то же время такой мониторинг активно используется бизнес-заказчиками компании, для того чтобы анализировать изменения в активности пользователей и успешность новых продуктов компании.

Универсальность подхода к мониторингу ГРВК и расчету метрик позволяет добавлять новые бизнес-метрики без модификации структуры БД, а только через внесение необходимой настройки в таблицу описания метрик и добавление SQL-запроса в процедуру сбора метрик. Все остальные элементы системы не требуют доработки.

Внедрение данного подхода позволяет повысить быстроту и надежность внесения изменений в систему, снизить затраты на создание системных и особенно бизнес-метрик, упростить и ускорить процесс обнаружения аномалий, сократить время простоев при плановой профилактике. Для созданных метрик можно настраивать пороги срабатывания триггеров и давать прогноз изменения бизнес-параметров на определенный период времени в будущем, используя известные математические модели прогнозирования [13].

Заключение

В результате выполненной работы создана универсальная структура данных и подход к мониторингу бизнес-метрик в различных областях применения. Предложенное решение не зависит от конкретных метрик, что позволяет унифицировать хранение в БД и мониторинг новых метрик с возможностью их динамического изменения без модификации структуры БД.

В статье показаны преимущества предложенного подхода и пример практической реализации бизнес-метрики «Количество загруженных файлов» в международной телекоммуникационной компании. Внедрение универсальной структуры БД позволило кардинально сократить затраты на организацию новых бизнес-метрик и ускорить оперативное обнаружение возможных аномалий в системе.

Предложенное решение универсально по отношению к используемой системе мониторинга, типу БД, программным средствам практической реализации, а также математическим моделям и подходам к прогнозированию вычислительных ресурсов ГРВК.

Список используемых источников

1. Distillery. URL: https://distillery.com (дата обращения 12.08.2020)
2. Dhyani A. Create a Custom Metric in Zabbix. 2016. URL: https://www.tothenew.com/blog/create-a-custom-metric-in-zabbix (дата обращения 12.08.2020)
3. Efimov V.V., Mescheryakov S.V., Shchemelinin D.A., Yakovlev K.A. Integration and Continuous Service Delivery in Globally Distributed Computing System // Университетский научный журнал. 2017. № 30. С. 13–20.
4. Gildeh D. What We Learnt Talking to 60 Companies about Monitoring // Dataloop.IO. 2014. URL: https://dataloopio.wordpress.com/2014/01/30/what-we-learnt-talking-to-60-companies-about-monitoring (дата обращения 12.08.2020)
5. Levey T. Introducing Real-Time Business Metrics // AppDynamics. 2013. URL: https://www.appdynamics.com/blog/news/introducing-real-time-business-metrics (дата обращения 12.08.2020)
6. Cliffe D. Monitoring Business Metrics and Refining Outage Response // PagerDuty. 2015. URL: https://www.pagerduty.com/blog/monitoring-business-metrics (дата обращения 12.08.2020)
7. Prometheus. Open-source Systems Monitoring and Alerting Toolkit. URL: https://prometheus.io (дата обращения 12.08.2020)
8. Nesvold H. Getting into Business with Prometheus. 2016. URL: https://schibsted.com/blog/business-with-prometheus (дата обращения 12.08.2020)
9. Saymon. URL: https://saymon.info/en-version (дата обращения 12.08.2020)
10. Gartner Magic Quadrant for EFSS (Enterprise File Sharing and Sync) // Content Collaboration Platforms. 2018. URL: https://www.gartner.com/en/documents/3881863/magic-quadrant-for-content-collaboration-platforms (дата обращения 12.08.2020)
11. Kucherova K., Mescheryakov S., Shchemelinin D. Prediction Experience and New Model // Proceedings of the 7th Annual International Zabbix Conference (Riga, Latvia, 15–17 September 2017). URL: http://www.zabbix.com/conf2017_agenda.php
12. Kucherova K.N., Mescheryakov S.V., Shchemelinin D.A. Using Predictive Monitoring Models in Cloud Computing Systems. Communications in Computer and Information Science // Proceedings of the 21st International Conference on Distributed Computer and Communication Networks (DCCN 2018, Moscow, Russia, 17–21 September 2018). Cham: Springer, 2018. Vol. 919. PP. 341–352. DOI:10.1007/978-3-319-99447-5_29
13.Mescheryakov S., Shchemelinin D., Efimov V. Adaptive control of cloud computing resources in the internet telecommunication multiservice system // Proceedings of the 6th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT-2014, St.-Petersburg, Russia, 6-8 October 2014). IEEE, 2014. PP. 287–293. DOI:10.1109/ICUMT.2014.7002117

* * *

Prediction of Cloud Computing Resources Based on the Open Source Monitoring System

K. Kucherova[1]

[1]Peter the Great Saint-Petersburg Polytechnic University,
St. Petersburg, 195251, Russian Federation

Article info
DOI:10.31854/1813-324X-2020-6-3-100-106
Received 05th August 2020
Accepted 22th September 2020

For citation: Kucherova K. Prediction of Cloud Computing Resources Based on the Open Source Monitoring System. *Proc. of Telecom. Universities.* 2020;6(3):100–106. (in Russ.) DOI:10.31854/1813-324X-2020-6-3-100-106

Abstract: *The paper describes the universal approach for monitoring the data storage of a globally distributed cloud computing system, which allows you to automate creation of new metrics in the system and predict their behavior for the end users. Since the existing monitoring software products provide built-in scheme only for system metrics like RAM, CPU, disk drives, network traffic, but don't offer solutions for business functions, IT companies have to design specialized database structure (DB). The data structure proposed in this paper for storing the monitoring statistics is universal and allows you to save resources when orginizing database monitoring on the scale of the GDCCS. The goal of the research is to develop a universal model for monitoring and forecasting of data storage in a globally distributed cloud computing system and its adequacy to real operating conditions.*

Keywords: *cloud computing system, database, monitoring, system metrics, prediction, correlation function.*

References

1. *Distillery*. Available from: https://distillery.com [Accessed 12th August 2020]
2. Dhyani A. *Create a Custom Metric in Zabbix*. 2016. Available from: https://www.tothenew.com/blog/create-a-custom-metric-in-zabbix [Accessed 12th August 2020]
3. Efimov V.V., Mescheryakov S.V., Shchemelinin D.A., Yakovlev K.A. Integration and Continuous Service Delivery in Globally Distributed Computing System. *Humanities and Science University Journal*. 2017;30:13–20.
4. Gildeh D. What We Learnt Talking to 60 Companies about Monitoring. *Dataloop.IO.* 2014. Available from: https://dataloopio.wordpress.com/2014/01/30/what-we-learnt-talking-to-60-companies-about-monitoring [Accessed 12th August 2020]
5. Levey T. Introducing Real-Time Business Metrics. *AppDynamics*. 2013. Available from: https://www.appdynamics.com/blog/news/introducing-real-time-business-metrics [Accessed 12th August 2020]
6. Cliffe D. Monitoring Business Metrics and Refining Outage Response. *PagerDuty*. 2015. Available from: https://www.pagerduty.com/blog/monitoring-business-metrics [Accessed 12th August 2020]
7. *Prometheus*. Open-source Systems Monitoring and Alerting Toolkit. Available from: https://prometheus.io [Accessed 12th August 2020]
8. Nesvold H. *Getting into Business with Prometheus*. 2016. Available from: https://schibsted.com/blog/business-with-prometheus [Accessed 12th August 2020]
9. *Saymon*. Available from: https://saymon.info/en-version [Accessed 12th August 2020]
10. *Content Collaboration Platforms*. Gartner Magic Quadrant for EFSS (Enterprise File Sharing and Sync). 2018. Available from: https://www.gartner.com/en/documents/3881863/magic-quadrant-for-content-collaboration-platforms [Accessed 12th August 2020]
11. Kucherova K., Mescheryakov S., Shchemelinin D. Prediction Experience and New Model. *Proceedings of the 7th Annual International Zabbix Conference, 15–17 September 2017, Riga, Latvia*. Available from: http://www.zabbix.com/conf2017_agenda.php [Accessed 12th August 2020]
12. Kucherova K.N., Mescheryakov S.V., Shchemelinin D.A. Using Predictive Monitoring Models in Cloud Computing Systems. Communications in Computer and Information Science. *Proceedings of the 21st International Conference on Distributed Computer and Communication Networks. DCCN 2018, 17–21 September 2018, Moscow, Russia*. Cham: Springer; 2018. vol.919. p.341–352. DOI:10.1007/978-3-319-99447-5_29
13. Mescheryakov S., Shchemelinin D., Efimov V. Adaptive control of cloud computing resources in the internet telecommunication multiservice system. *Proceedings of the 6th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops, ICUMT-2014, 6-8 October 2014, St.-Petersburg, Russia.* IEEE; 2014. p.287–293. DOI:10.1109/ICUMT.2014.7002117

Сведения об авторе:

КУЧЕРОВА Кристина Николаевна | аспирант высшей школы креативной индустрии и дизайна института машиностроения, материалов и транспорта Санкт-Петербургского политехнического университета Петра Великого, kristina.mylife@gmail.com

РОССИЙСКАЯ НЕДЕЛЯ ВЫСОКИХ ТЕХНОЛОГИЙ

Минкомсвязь России

МИНПРОМТОРГ РОССИИ

Федеральное агентство связи (РОССВЯЗЬ)

ТПП РФ

Expo Rating

СВЯЗЬ

Информационные и коммуникационные технологии

2–6 ноября 2020

ВНИМАНИЕ!
НОВЫЕ ДАТЫ!

32-я международная выставка

Организатор:

ЭКСПОЦЕНТР

При поддержке:
- Министерства цифрового развития, связи и массовых коммуникаций РФ
- Министерства промышленности и торговли РФ
- Федерального агентства связи (РОССВЯЗЬ)

Под патронатом ТПП РФ

Россия, Москва, ЦВК «ЭКСПОЦЕНТР»

www.sviaz-expo.ru

12+ Реклама

Выходные данные

Дизайн обложки – ООО «Комильфо»

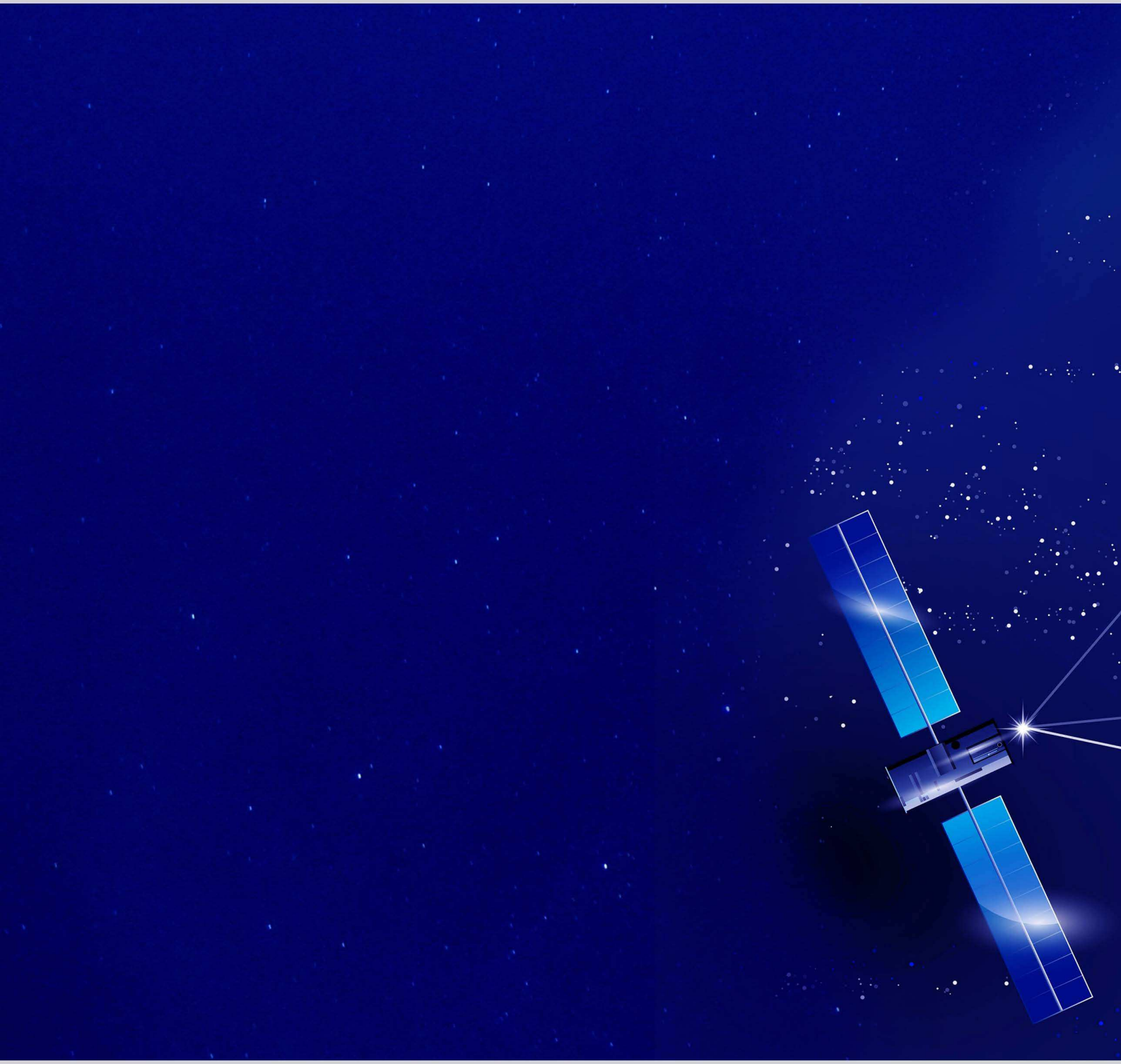
План издания научной литературы 2020 г., п. 5

Дата выхода в свет	Усл.-печ. л.	Формат	Тираж	Заказ	Свободная цена
29.09.2020	13,5	$60\times84_{1/8}$	1000 экз.	№ 1147	

Ответственный редактор **Татарникова И.М.**
Выпускающий редактор **Яшугин Д.Н**

Отпечатано в СПбГУТ
193232, Санкт-Петербург, пр. Большевиков, 22/1

Учредитель и издатель:
Федеральное государственное бюджетное образовательное учреждение
высшего образования "Санкт-Петербургский государственный университет
телекоммуникаций им. проф. М.А. Бонч-Бруевича"
E-mail: tuzs@spbgut.ru Web: tuzs.sut.ru VK: vk.com/spbtuzs



Подписной индекс по каталогу "Издания органов НТИ" Агентства "Роспечать" – 59983