ISSN 2225-5435

Вестник УрФО.

БЕЗОПАСНОСТЬ В ИНФОРМАЦИОННОЙ СФЕРЕ

№ 1(35) / 2020

УЧРЕДИТЕЛИ

ФГАОУ ВО «ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

ООО «ЮЖНО-УРАЛЬСКИЙ ЮРИДИЧЕСКИЙ ВЕСТНИК»

Подписной индекс 73852 в каталоге «Почта России»

Журнал зарегистрирован Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций.

Свидетельство ПИ № ФС77-65765 от 20.05.2016

Издатель: **ООО «Южно-Уральский юридический вестник»**

Адрес редакции и издателя: Россия, 454080, г. Челябинск, пр. Ленина, д. 76.
Тел./факс (351) 267-97-01.

Электронная версия журнала в Интернете:
www.info-secur.ru,
e-mail: urvest@mail.ru

16+

Journal of the Ural Federal District. Information security

ISSN 2225-5435

№ 1(35) / 2020

FOUNDER

SOUTH URAL STATE UNIVERSITY

SOUTH URAL LEGAL NEWSLETTER

CHAIRMAN OF THE EDITORIAL BOARD

CHUVARDIN O. P.,

Head of Department Federal Service for Technical and Export Control of Russia for the Urals Federal District

CHIEF EDITOR

SOKOLOV A.N.,

Ph.D., Associate Professor, Head of Department "Information Protection", South Ural State University (National Research University) (Chelyabinsk city)

PRODUCING EDITOR

SOGRIN E. K.

LAYOUT

SHRABER A. E.

PROOFREADING

FEDOROV V. S.

Subscription index 73852

in the «Russian Post» catalog

The journal is registered by the Federal service in the field of communication, information technology and mass communications.

Certificate
PI No. ФС77-65765 dd. 05/20/2016

Publisher: OOO « South Ural Legal Newsletter»

Editorial and publisher address: Russia, 454080, Chelyabinsk, Lenin Avenue, 76
Phone / fax (351) 267-97-01.

Electronic version of the magazine in the Internet:

www.info-secur.ru,
e-mail: urvest@mail.ru

16+

В НОМЕРЕ

ИССЛЕДОВАНИЕ И ПРОЕКТИРОВАНИЕ ТЕХНИЧЕСКИХ СРЕДСТВ

МЕТОДЫ АНАЛИЗА ДАННЫХ

ОРГАНИЗАЦИОННО-ТЕХНИЧЕСКАЯ И ПРАВОВАЯ ЗАЩИТА ИНФОРМАЦИИ

АКТУАЛЬНЫЕ ПРОБЛЕМЫ КИБЕРБЕЗОПАСНОСТИ

RESEARCH AND DESIGN OF TECHNICAL FACILITIES

METHODS OF DATA ANALYSIS

ORGANIZATIONAL, TECHNICAL AND LEGAL PROTECTION OF INFORMATION

TOPICAL PROBLEMS OF CYBERSECURITY

УДК 004.056.5 Вестник УрФО № 1(35) / 2020, с. 5–11

Субботин С. Д., Поршнев С. В., Пономарева О. А. DOI: 10.14529/secur200101

ИССЛЕДОВАНИЕ ПОВЕДЕНИЯ ИНФОРМАТИВНЫХ СИГНАЛОВ ОТ НАКОПИТЕЛЯ НА ЖЕСТКИХ МАГНИТНЫХ ДИСКАХ ПРИ ОБРАБОТКЕ ИНФОРМАЦИИ В РАЗНЫХ ОБЛАСТЯХ ДИСКА

В статье демонстрируется изменение спектра информативных сигналов во время моделирования работы (чтения или записи информации) накопителя на жестких магнитных дисках в разных его областях с помощью специализированного программного обеспечения из состава программно-аппаратных комплексов для проведения инженерных исследований и исследований на сверхнормативные побочные электромагнитные излучения и наводки с целью обнаружения возможных частот информативных сигналов (по которым вероятна утечка обрабатываемой информации в том числе ограниченного распространения).

Ключевые слова: *специальные исследования, СИ, накопитель на жестких магнитных дисках, НЖМД, информативные сигналы, электрические сигналы, ПЭМИН, тест-программа, электромагнитное излучение, ЭМИ, защита информации, канал утечки информации, интерфейс SATA.*

THE STUDY OF THE BEHAVIOR OF INFORMATIVE SIGNALS FROM THE ON HARD DISK DRIVE WITH THE PROCESSING OF INFORMATION IN DIFFERENT AREAS OF THE DISK

The article demonstrates the change in the spectrum of informative signals during the simulation of work (reading or writing information) hard disk drive in its different areas with the help of specialized software from the composition of software and hardware complexes for engineering research and research transient electromagnetic pulse emanation standard in order to detect the possible frequencies of informative signals (which are likely to leak the processed information including of limited distribution).

Keywords: *special study, hard disk drive, HDD, informative signals, electrical signal, TEMPEST, test program, electromagnetic radiation, EM radiation, information protection, information leakage channel, SATA interface.*

Одним из важных элементов основных технических средств и систем (далее – ОТСС) является накопитель на жестких магнитных дисках (далее – НЖМД), которым не стоит пренебрегать при проведении специальных исследований (далее – СИ). Классический IDE интерфейс НЖМД – параллельный, а интерфейсы с параллельным кодированием и разрядностью выше 16 как опасные по каналу ПЭМИН уже рассматривать не имеет смысла, также вид данного интерфейса все реже используется, а вот интерфейс SATA в различных своих вариантах является последовательным. Информативные сигналы от SATA интерфейса выявляются по-разному, часто их за пределами корпуса системного блока обнаружить просто не удается [1]. Но это не означает, что если специалисту не удалось их обнаружить, то их нет вообще, ведь если сегодня существующими техническими средствами потенциальный противник не смог осуществить разведку, цель которой определить – что передавалось в конкретный момент времени, ноль или единица, завтра усовершенствовав свои навыки и средства разведки он сможет реализовать съем необходимой информации.

Цель статьи установить изменение спектра информативных сигналов во время моделирования работы НЖМД в разных его областях тест-сигналами, создаваемыми специализированным программным обеспечением (далее – тест-программы) из состава комплексов для проведения инженерных исследований и исследований на сверхнормативные побочные электромагнитные излучения и наводки в конкретной модели НЖМД, представленой в табл. 1 и 3 [2].

Детальные характеристики исследуемого НЖМД из официальной документации представлены в табл. 2 и 3 [3].

Поиск информативных сигналов от исследуемого НЖМД осуществлялся техническими средствами, представленными в табл. 3.

В качестве тест-сигналов использовались тест-программы: «макет специального теста дисковых накопителей» и «СИГУРД-Test» с настроенным сплошным спектром тест-сигнала и параметром заполнения двоичным кодом «1010101» для режимов чтения и записи (см. рис. 1).

Перед проведением СИ исследуемый

Таблица 1

Данные об объекте исследования

Наименование исследуемого устройства	Марка, модель	Серийный номер	Интерфейс
НЖМД	Seagate Barracuda 7200.10 ST3250410AS (250 Gb)	9RY0B525	SATA 2

Таблица 2

Подробные характеристики Seagate Barracuda 7200.10

Количество дисков	1
Количество считывающих головок	2
Гарантированное количество секторов	488 397 168
Количество секторов на дорожке	63
Количество дорожек на всем диске	488 397 168 / 63 = 7 752 336
Количество цилиндров	7 752 336 / 2 = 3 876 168
Скорость вращения шпинделя	7200 об/мин

Таблица 3

Технические средства для проведения исследования

Наименование средств и программ для измерений	Тип	Заявленный производителем диапазон частот, МГц	Фактически использованный диапазон частот, МГц
Программно определяемая радиосистема (SDR)	USRP B210	70 - 6000	34,5 - 6001 (расширен программой USRP Spectrum Scanner)
Антенна широкополосная всенаправленная	Gabil GRA-3000M	2 - 3000	34,5 - 3000
Антенна широкополосная всенаправленная	Ultra-Base Antenna 08-ANT-0861	25 - 6000	3001 - 6000
Программное обеспечение	USRP Spectrum Scanner (FFT)	-	-

НЖМД подвергся полному форматированию в файловой системе NTFS со стандартным размером кластера в 4096 байт. Далее, применив тест-программы, в списке доступных носителей был выбран исследуемый НЖМД. После нажатия кнопки «Запись» отправляется команда считывающим головкам НЖМД на запись тест-файла в ближайшую свободную область диска. Таким образом было записано два тест-файла программой «макет специального теста дисковых накопителей» и два тест-файла программой «СИГУРД-Test» с разными именами. Тест-программы представлены на рис. 1. Затем программой WinHex каждый второй записанный тест-файл был перенесен в другую область диска переназначением кластера для изменения положения считывающих головок НЖМД при обращении к этим файлам.

После проведения СИ ПЭМИН, результат которого приведен в табл. 4 видно, что спектр выявленных информативных сигналов от исследуемого устройства отличается в зависимости от расположения тест-файла на диске (от положения считывающих головок).

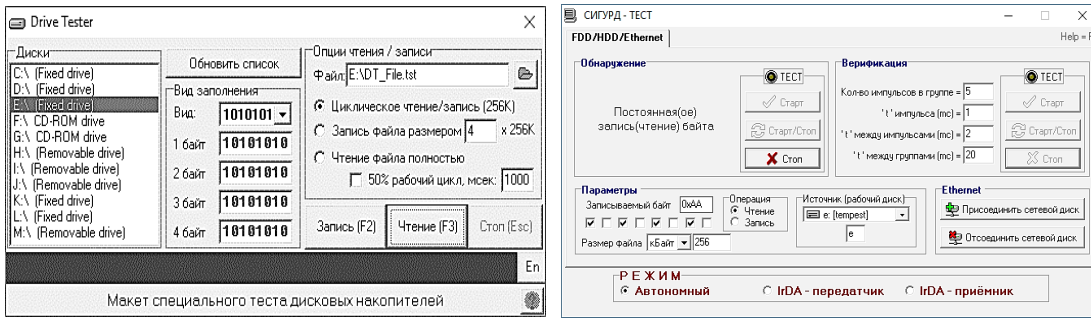


Рис. 1. Интерфейс тест-программ: «Макет специального теста дисковых накопителей» и «СИГУРД-Test»

Таблица 4

Результат исследования НЖМД

№ п/п	Частота информативного сигнала, МГц	Тест режим	Информативный сигнал при тест-файле в первой области НЖМД	Информативный сигнал при тест-файле во второй области НЖМД
1.	247,165	Запись	Присутствует	Присутствует
2.	254,323	Чтение или Запись	Присутствует	Отсутствует
3.	255,255	Чтение	Отсутствует	Присутствует
4.	300,961	Чтение	Присутствует	Присутствует
5.	339,042	Чтение	Присутствует	Присутствует
6.	508,640	Чтение или Запись	Присутствует	Отсутствует
7.	529,836	Чтение или Запись	Присутствует	Отсутствует
8.	1729,422	Запись	Присутствует	Присутствует
9.	1929,700	Запись	Присутствует	Присутствует

Также был зафиксирован электрический сигнал от НЖМД на тактовой частоте 3000,108 МГц интерфейса SATA 2 и частоте 6000,213 МГц, не имеющий признаков информативности от тест-программ, но возникающий при включении исследуемого НЖМД, поэтому данные сигналы во внимание не берутся.

Поиск частот информативных сигналов исследуемого НЖМД осуществлялся в ручном режиме с полосой пропускания измерительного приемника от 1до 10 МГц для диапазона от 34,5 МГц до 3000 МГц и 20 МГц от 3000 МГц до 6000 МГц, рис. 2.

На процесс выявления сигналов ограничений не существует, поэтому антенна размещалась в месте максимального излучения, рис. 3. Поставив разведчика в лучшие условия, но худшие для нас, мы уменьшаем вероятность некачественной защиты обнаруженного информативного сигнала.

Существующие нормативно-методические документы по проведению СИ требуют осуществлять поиск информативных сигналов с разных сторон исследуемого элемента ОТСС, а поскольку сам НЖМД небольшого размера и внутри механические части отвечающие за процесс записи и чтения, являются миниатюрными, то расположение измерительной антенны вряд ли повлияет на результат исследования.

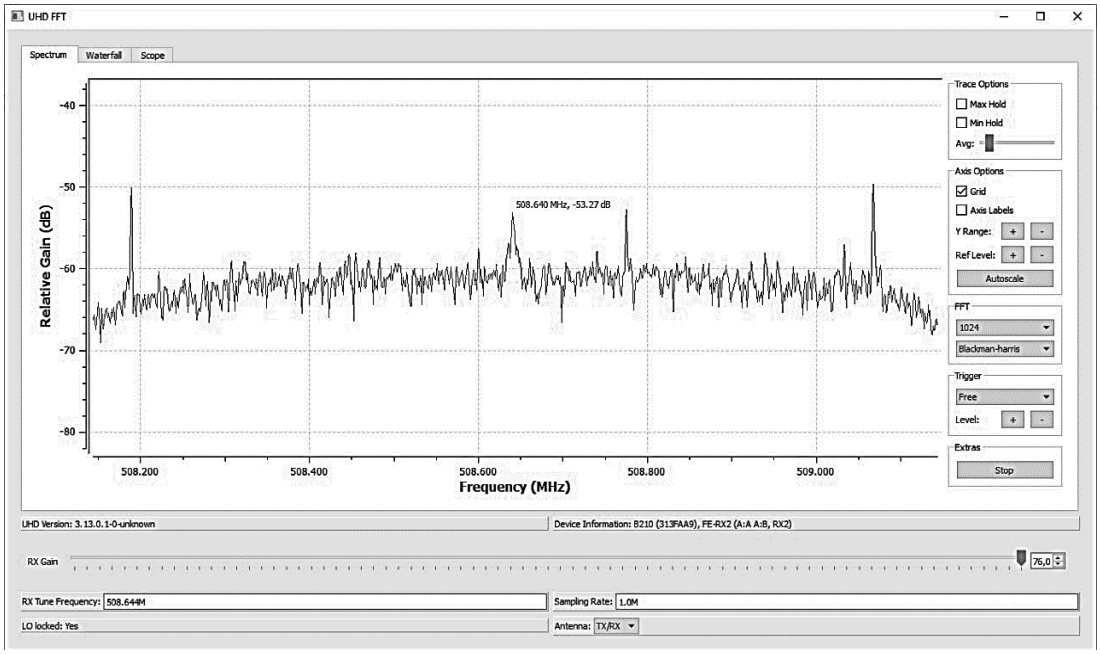


Рис. 2. Интерфейс программы USRP Spectrum Scanner (FFT)

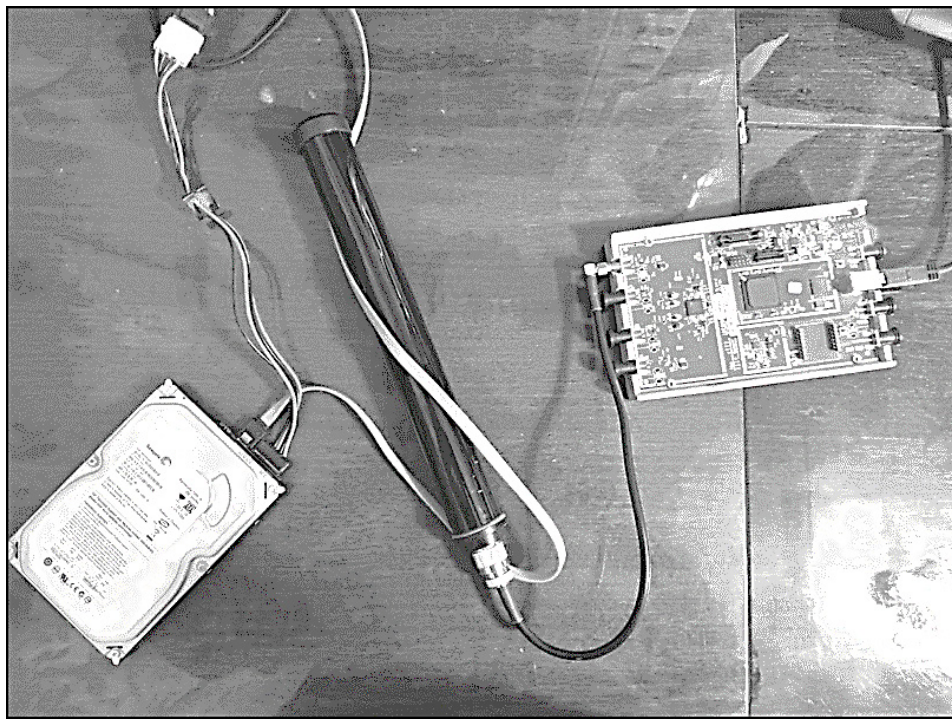


Рис. 3. Проведение поиска информативных сигналов от НЖМД

Рассмотрев принцип работы НЖМД, учитывая, что считывающие головки привязаны к месту обработки информации на диске, можно сделать вывод, что видимое изменение спектра связано с геометрическим положением считывающих головок внутри конструкции НЖМД, а рассмотренные тест-программы, моделирующие работу НЖМД, не учитывают данную особенность. Объяснение изменения спектра сигнала относительно расположения считывающих головок НЖМД при обработке информации заключается в различном переотражении сигнала из-за геометрии самого корпуса НЖМД и его элементов внутри, а также в изменении угловой скорости и длины окружности конкретной дорожки диска.

Почему необходимо проводить СИ НЖМД, моделируя обработку информации в разных областях диска? Потому что файлы со временем имеют свойство фрагментироваться по всей области диска. Либо, когда различные действия пользователя связанные с обработкой информации на компьютере провоцируют появление новых частот информативных сигналов, приводя актуальный протокол технического контроля в несоответствие текущему состоянию системы из-за многократного освобождения или заполнения неопределенного пространства НЖМД за сессию.

Для эффективного проведения СИ НЖМД предлагается разработка тест-программы, с возможностью учёта количества дисков внутри НЖМД, а также положение считывающих головок при обработке информации, позволяя специалисту записывать тест-файл в разные области диска независимо от текущего объема НЖМД.

Литература

1. Кондратьев А.В. Техническая защита информации. Практика работ по оценке основных каналов утечки. – М.: Горячая линия – Телеком, 2016. — 304 с. – ISBN 978-5-9912-0574-0.

2. Субботин С.Д., Поршнев С.В., Пономарева О.А. Исследование эффективности тест-сигналов, создаваемых специализированным программным обеспечением при проведении специальных исследований накопителей на жестких магнитных дисках для выявления технического канала утечки информации ПЭМИН // Сборник трудов XVIII Всероссийской научно-практической конференции студентов, аспирантов и молодых ученых «Безопасность информационного пространства». – Магнитогорск : Изд-во Магнитогорск. гос. техн. ун-та им. Г.И. Носова, 2019. – С. 53-58 – ISBN 978-5-9967-1764-4.

3. Product Manual Barracuda 7200.10 Serial ATA. [Электронный ресурс] Seagate.com. URL: https://www.seagate.com/staticfiles/support/disc/manuals/deskt- op/Barracuda%207200.10/100402371k.pdf.

References

1. Kondrat'ev A.V. Tehnicheskaja zashhita informacii. Praktika rabot po ocenke osnovnyh kanalov utechki. [Technical protection of information. Practice of works on estimation of the main channels of leakage]. – M.: Gorjachaja linija – Telekom, 2016. 304 s. – ISBN 978-5-9912-0574-0.

2. Subbotin S.D., Porshnev S.V., Ponomareva O.A. Issledovanie jeffektivnosti test-signalov, sozdavaemyh specializirovannym programmnym obespecheniem pri provedenii special'nyh issledovanij nakopitelej na zhjostkih magnitnyh diskah dlja vyjavlenija tehnicheskogo kanala utechki informacii PJEMIN [Study of the effectiveness of test signals generated by specialized software during special studies of hard disk drives to identify the technical potential of information leakage TEMPEST] // Sbornik trudov XVIII Vseros-siyskoy nauchno-prakticheskoy konferentsii studentov, aspirantov i molodykh uchenykh «Bezopasnost' informatsionnogo prostranstva» [Proceedings of the XVIII All-Russian Scientific and Practical Conference of Students, Graduate Students and Young Scientists «Information Space Security»]. Magnitogorsk, Izd-vo Magnitogorsk. gos. tekhn. un-ta im. G.I. Nosova, 2019, pp. 302-306. ISBN 978-5-9967-1764-4.

3. Product Manual Barracuda 7200.10 Serial ATA. [Электронный ресурс] Seagate.com. URL: https://www.seagate.com/staticfiles/support/disc/manuals/deskt- op/Barracuda%207200.10/100402371k.pdf.

СУББОТИН Станислав Дмитриевич, аспирант Института радиоэлектроники и информационных технологий - радиотехнический факультет Уральского федерального университета им. Первого Президента России Б.Н. Ельцина. 620002, г. Екатеринбург, ул. Мира, 19. E-mail: stwantod@gmail.com.

ПОШНЕВ Сергей Владимирович, доктор технических наук, профессор, директор Учебно-научного центра «Информационная безопасность» Уральского федерального университета им. Первого Президента России Б.Н. Ельцина. 620002, г. Екатеринбург, ул. Мира, 19. E-mail: s.v.porshnev@urfu.ru.

ПОНОМАРЕВА Ольга Алексеевна, старший преподаватель Института радиоэлектроники и информационных технологий - радиотехнический факультет Уральского федерального университета им. Первого Президента России Б.Н. Ельцина. 620002, г. Екатеринбург, ул. Мира, 19. E-mail: o.a.ponomareva@urfu.ru.

SUBBOTIN Stanislav Dmitrievich, Postgraduate of Institute of Radio electronics and Information

Technologies, Ural Federal University named after first President of Russia B.N. Yeltsin. 620002, Yekaterinburg, Mira str., 19. E-mail: stwantod@gmail.com.

POSHNEV Sergej Vladimirovich, Doctor of Technical Sciences, Full Professor, Head of Unit, Training and Research Center «Information Security», Ural Federal University named after first President of Russia B.N. Yeltsin. 620002, Yekaterinburg, Mira str., 19. E-mail: s.v.porshnev@urfu.ru.

PONOMAREVA Olga Alekseevna, Senior Lecturer of Institute of Radio electronics and Information Technologies, Ural Federal University named after first President of Russia B.N. Yeltsin. 620002, Yekaterinburg, Mira str., 19. E-mail: o.a.ponomareva@urfu.ru.

УДК 004.056 Вестник УрФО № 1(35) / 2020, с. 12–23

Гайдамакин Н. А., Синадский Н. И., Сушков П. В.

DOI: 10.14529/secur200102

КОМПЛЕКСНЫЙ ИМИТАЦИОННО-СТАТИСТИЧЕСКИЙ МЕТОД СИНТЕЗА МАССИВОВ УСЛОВНО-РЕАЛЬНЫХ ДАННЫХ НА ОСНОВЕ СТРУКТУРНО-ПАРАМЕТРИЧЕСКОЙ МОДЕЛИ ВЗАИМОДЕЙСТВИЯ ПОЛЬЗОВАТЕЛЕЙ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННЫХ СЕРВИСОВ

В статье представлено решение задачи синтеза учебных заданий и массивов данных при организации компьютерного полигона для проведения практических занятий по расследованию инцидентов информационной безопасности. Предложены два основных этапа синтеза фоновой и ситуационной составляющих массивов условно-реальных данных на основе структурно-параметрической модели взаимодействия пользователей информационно-телекоммуникационных сервисов: формирование статических и динамических компонентов.

Статические компоненты синтезируются на основе метода формирования структуры социальных графов, использующего композицию моделей построения сложных сетей с различными структурными параметрами: для сервиса мобильной связи применяется модель Ваттса-Строгатца, для сервиса социальных сетей – модель Барабаши-Альберт. Для сохранения взаимосвязи между пользователями в различных сервисах предложен метод определения наибольшей общей части социальных

графов, основанный на взаимной дифференциации вершин и выделении частичного изоморфизма сравниваемых графов. При формировании атрибутов вершин применяется метод поиска социальных групп (семей), основанный на алгоритме Брона-Кербоша по поиску клики заданного размера в графе.

Для синтеза динамических компонентов массивов данных, описывающих совершение коммуникационных событий, используется математический аппарат цветных сетей Петри. Событие взаимодействия в информационно-телекоммуникационных сервисах представляется в виде метки сети Петри, которая содержит необходимый набор параметров, зависящий от типа сервиса. Для формирования начальной разметки сети Петри предложено использовать структурные, событийные, социальные и временные статистические характеристики реальных информационно-телекоммуникационных сервисов.

Ключевые слова: *массив условно-реальных данных, информационно-телекоммуникационные сервисы, модели сложных сетей, цветная сеть Петри, учебный компьютерный полигон.*

Gaidamakin N. A., Sinadsky N. I., Sushkov P. V.

COMPLEX SIMULATION-STATISTICAL METHOD FOR SYNTHESIZING CONDITIONALLY REAL DATA ARRAYS BASED ON A STRUCTURAL-PARAMETRIC MODEL OF INTERACTION BETWEEN USERS OF INFORMATION AND TELECOMMUNICATION SERVICES

The article presents a solution to the problem of synthesizing training tasks and data arrays when organizing a computer training platform for conducting practical exercises to investigate information security incidents. Two main stages of the synthesis of the background and situational components of the conditionally real data arrays based on the structural-parametric model of interaction between users of information and telecommunication services are proposed: the formation of static and dynamic components.

Static components are synthesized based on the method of forming the structure of social graphs using a composition of models for constructing complex networks with various structural parameters: the Watts-Strogatz model is used for a mobile communication service, and the Barabashi-Albert model is used for a social network service. To preserve the relationship between users in various services, a method is proposed for determining the largest common part of social graphs, based on the mutual differentiation of vertices and the allocation of a partial isomorphism of the compared graphs. When generating vertex attributes, the method of searching for social groups (families) is used, based on the Bron-Kerbosch algorithm for finding a click of a given size in a graph.

For the synthesis of dynamic components of data arrays that describe the performance of communication events, the mathematical apparatus of color Petri nets is used. An interaction event in information and telecommunication services is represented in the form of a Petri net label, which contains the necessary set of parameters, depending on the type of service. It is proposed to use the structural, event, social and temporal statistical characteristics of real information and telecommunication services to form the initial marking of the Petri net.

Keywords: *array of conditionally real data, information and telecommunication services, models of complex networks, color Petri net, training computer training platform.*

Вступление в силу Федерального закона «О безопасности критической информационной инфраструктуры Российской Федерации» свидетельствует об актуальности и значимости решения задачи по обнаружению, предупреждению и ликвидации последствий компьютерных атак и реагирования на компьютерные инциденты. Условием качественного решения данных задач является практико-ориентированная профессиональная подготовка соответствующих специалистов.

Для организации и проведения практических занятий по расследованию инцидентов информационной безопасности в сетях документальной электросвязи и сети Интернет как на потоках магистратуры по направлению «Информационная безопасность», так и на потоках специалитета «Информационная безопасность телекоммуникационных систем» и «Информационно-аналитические системы безопасности» необходимо создание учебного компьютерного полигона, оснащенного современными образцами информационно-аналитических систем безопасности (далее – ИАСБ), такими как IBM I21, МФИ СОФТ «Январь»2, Lampyre3, Gephi4 и др.

ИАСБ – это аппаратно-программные комплексы для проведения поисково-аналитической работы, имеющие возможность накапливать и анализировать данные о взаимодействии пользователей информационно-телекоммуникационных сервисов (далее – ИТ-сервисов). Однако подключение учебных ИАСБ к действующему оборудованию операторов связи, являющемуся источником информации о взаимодействии пользователей ИТ-сервисов, невозможно в соответствии со ст. 64 Федерального закона «О связи». Также отсутствует возможность применения настоящих массивов биллинговой информации в силу того, что такие массивы содержат персональные данные пользователей, а доступ к ним ограничен законодательно.

Обзор литературы позволяет сделать вывод об отсутствии готовых методов и алгоритмов генерации массивов данных, отражающих взаимодействие пользователей ИТ-сервисов.

Для решения данной проблемы при создании учебного компьютерного полигона по расследованию инцидентов информационной безопасности в учебно-научном центре «Информационная безопасность» ИРИТ-РтФ УрФУ им. первого Президента России Б.Н. Ельцина разработано программное обеспечение (далее – ПО). Данное ПО работает в соответствии с созданной структурно-параметрической моделью взаимодействия пользователей в ИТ-сервисах и позволяет синтезировать массивы условно-реальных данных. Файлы, содержащие сгенерированные массивы, в дальнейшем загружаются в базу данных ИАСБ для проведения практических занятий по решению поисково-аналитических задач. На текущем этапе разработки ПО способно генерировать массивы условно-реальных данных о взаимодействии пользователей социальных сетей и сетей мобильной связи.

[1] http://www.ibm.com/software/ products/ru/analysts-notebook

[2] http://www.mfisoft.ru/direction/sorm/sorm-3/

[3] http://www.lampyre.io

[4] http://www.gephi.org

Структура массивов условно-реальных данных

Данные, подлежащие анализу, накапливаются в ИАСБ в формате совокупности таблиц взаимодействия пользователей. В зависимости от вида ИАСБ и типа ИТ-сервиса таблицы, имеющие различные поля хранимой информации, описываются множеством записей вида:

$$D=\{d_1, d_2, \ldots, d_n\},$$

где d_n – элемент таблицы (строка, запись), описывающий одно коммуникационное событие. Для сервиса мобильной связи строка параметров, хранящихся в памяти ИАСБ, может иметь следующий вид [1]:

$$d_{Mobile} \in D_{Mobile} = \{AbonentIMSI, AbonentIMEI, AbonentPhone, LAC, CellID, Time, CallDuration, BillingType, PhoneB\},$$

где *AbonentIMSI* — идентификатор абонента, *AbonentIMEI* — серийный номер устройства абонента, *AbonentPhone* — номер телефона абонента, *BillTime* — время фиксации соединения, *CallDuration* — продолжительность соединения в секундах, *BillingType* — тип соединения, *LAC* — код локальной зоны, *CellId* — идентификатор соты, *PhoneB* — номер телефона принимающей соединение стороны.

Для сервиса социальных сетей – вид:

$$dS_{ocial} \in D_{Social} = \{AbonentPhone, AbonentUID, AbonentIP, AbonentLogin, Time, Type, UidB\},$$

где *AbonentPhone* — номер телефона пользователя, *AbonentUID* — идентификатор учетной записи пользователя, *AbonentIP* — IP-адрес пользователя, *AbonentLogin* — логин Интернет-соединения пользователя, *Time* — время фиксации события, *Type* — тип события, *UidB* — идентификатор учетной записи второго пользователя, участвующего в событии.

Для решения поисково-аналитических задач в структуре массива данных должны быть представлены фоновые и ситуационные компоненты [1]:

$$D = D_{task} \cup D_{feed},$$

где D_{task} – ситуационные задачи и D_{feed} – фоновые события.

Ситуационная задача предназначена для последующего выявления взаимодействия пользователей при решении комплексной учебной поисково-аналитической задачи, описывается преподавателем через интерфейс разработанного ПО. Фоновый массив данных имитирует социальную активность в целом и содержит события пользователей, подчиняющиеся некоторым статистическим законам реальных ИТ-сервисов.

Синтез фонового массива данных основан на комплексном имитационно-статистическом методе, включающем последовательное формирование статических и динамических компонентов с использованием различных математических подходов:

1. На первом этапе происходит формирование статических компонентов записей d_{Mobile} и d_{Social}, к которым относятся персональные идентификаторы пользователей (*AbonentIMSI, AbonentIMEI, AbonentPhone, AbonentUID, AbonentIP, AbonentLogin*) и структура их взаимосвязей внутри ИТ-сервиса (в том числе, информация о социальных группах). Для создания структурной основы, которая описывает существование социальных связей между пользователями, используется композиция существующих моделей построения сложных сетей на основе статистических распределений структурных параметров сервисов мобильной связи и социальных сетей.

2. На втором этапе на основе существующих социальных связей пользователей происходит генерация динамических компонентов коммуникационного события (тип, участники, место, время начала и продолжительность). Для формирования различных полей записей d_{Mobile} и d_{Social} требуются статистические распределения определенных параметров, которые можно разбить на категории:

– событийные (поля *BillingType, Type*);

– социальные (поля *PhoneB, UIDB*);

– пространственные (поля *LAC, CellID*);

– временные (поля *Time, CallDuration*).

Задача синтеза динамической составляющей фонового массива данных, предполагающая наличие большого количества активных объектов с отчетливо выраженным индивидуальным поведением, относится к категории синтеза сложных систем. Одним из подходов к синтезу сложных систем является агентное моделирование, позволяющее учесть структуру и взаимодействие пользователей ИТ-сервисов. В качестве общепринятого математического аппарата решения задач агентного имитационного моделирования применяются цветные сети Петри (далее – ЦСП).

Формирование статических компонентов массивов условно-реальных данных

ИТ-сервисы целесообразно рассматривать в виде социальных графов. В общем случае структура социального графа представляется в виде $G=(U, E)$, где U — множество вершин графа. Обозначим через G_M и G_S социальные графы ИТ-сервисов мобильной связи и социальных сетей соответственно. Ситуационная задача t описывается шаблоном взаимодействия пользователей $G_t = (U_t, E_t)$.

С целью создания основы для описания взаимодействия пользователей ИТ-сервисов предлагается использовать метод формирования статической структуры социальных графов G_M и G_S на основе композиции существующих моделей построения сложных сетей с учетом заданного шаблона взаимодействия пользователей G_t.

В рамках данного исследования наибольший интерес представляют три модели построения сложных сетей, позволяющие описывать взаимодействие между людьми:

– модель Эрдёша-Реньи (случайные графы) [2];

– модель Ваттса-Строгатца (сети тесного мира) [3];

– модель Барабаши-Альберт (сети предпочтительного присоединения) [4].

В таблице 1 представлены основные структурные свойства, которые являются определяющими при выборе модели для формирования структуры ИТ-сервиса.

сетей определены некоторые структурные свойства [4 ,5]:

– закон распределения степеней вершин близкий к Пуассоновскому;

– малая длина пути между вершинами;

– высокий коэффициент кластеризации;

– децентрализованная структура.

Построение структуры мобильных сетей начинается с некоторого фиксированного числа вершин, которые затем случайным образом связываются или меняют связи. Однако социальные сети носят открытый характер, что подразумевает рост благодаря непрерывному добавлению новых узлов. Начиная с небольшого ядра количество узлов увеличивается на протяжении всего времени жизни сети путем последующего добавления новых. Кроме того, большинство социальных сетей демонстрируют предпочтительное соединение, такое, что вероятность подключения к узлу зависит от его степени.

Указанные свойства определяют особенность социальных сетей по сравнению с сетями взаимодействия в реальном мире: закон распределения степеней вершин – показательный. Кроме того, показательному закону распределения соответствует сильноцентрализованная структура. Остальные свойства (малая длина пути между вершинами, высокий коэффициент кластеризации) остаются неизменными.

Таблица 1

Структурные свойства моделей

	Модель Эрдёша-Реньи	**Модель Ваттса-Строгатца**	**Модель Барабаши-Альберт**
Закон распределения степеней вершин	Пуассоновский	близкий к Пуассоновскому	показательный
Расстояние между вершинами	малое	малое	малое
Коэффициент кластеризации	низкий	высокий	средний
Структура	децентрализованная	децентрализованная	централизованная

Процесс формирования статических структур социальных графов целесообразно начинать с построения структуры сервиса мобильной связи, т.к. количество абонентов будет превышать количество пользователей социальных сетей.

Структура взаимосвязей абонентов ИТ-сервиса сотовой связи будет рассматриваться как сеть простого вербального общения людей в реальном мире без применения каких-либо технических средств, т.к. для данных

Анализ свойств моделей и реальных сетей позволяет сделать вывод, что для описания взаимодействия абонентов мобильной связи наилучшим образом подходит модель Ваттса-Строгатца. С другой стороны, активность между пользователями социальных сетей имеет отличительные особенности, наиболее полно отраженные в модели Барабаши-Альберт.

Процесс формирования структуры социального графа G_M начинается с генерации регу-

лярной решетки со степенью вершин *K*. Затем происходит выбор случайным образом соседних вершин в количестве $|U_t|$, с распределением значений вершин и ребер в соответствии с U_t и E_t. На последнем этапе выполняется перераспределение каждого ребра с вероятностью *p* на случайную вершину. Назначенные в соответствии с E_t ребра остаются неизменными.

Процесс формирования структуры социального графа G_S начинается с генерации случайного графа с количеством вершин m_0. Данный граф выступает в роли начального ядра будущего социального графа. Случайный граф строится в соответствии с моделью Эрдёша-Реньи. Исходными данными на этом этапе являются количество вершин m_0 и вероятность p_0, с которой между двумя произвольными вершинами образуется ребро.

После создания структурного ядра графа происходит последовательное добавление вершин в количестве m_{max}, определенном изначально. За один шаг создается одна вершина. Количество ребер, с которым данная вершина добавляется, зависит от коэффициента *C*, значение которого определено заранее и остается постоянным. Данный коэффициент призван снизить разрежённость формируемого графа, что как следствие позволит увеличить коэффициент кластеризации.

В процессе формирования атрибутивных параметров вершин социальных графов возникает необходимость нахождения «семей», т.е. групп пользователей, объединенных в полносвязный граф. Такая задача существует и в теории графов и носит название «Задача о клике». Кликой в неориентированном графе называется подмножество вершин, каждые две из которых соединены ребром графа. Иными словами, это полный подграф первоначального графа. Задача о клике существует в двух вариантах: в задаче распознавания требуется определить, существует ли в заданном графе *G* клика размера *k*, в то время как в вычислительном варианте требуется найти в заданном графе *G* клику максимального размера.

Алгоритм Брона — Кербоша — метод ветвей и границ для поиска всех клик (а также максимальных по включению независимых множеств вершин) неориентированного графа [6]. Алгоритм использует тот факт, что всякая клика в графе является его максимальным по включению полным подграфом. Начиная с одиночной вершины (образующей полный подграф), алгоритм на каждом шаге пытается увеличить уже построенный полный подграф, добавляя в него вершины из множества кандидатов. Высокая скорость обеспечивается отсечением при переборе вариантов, не приводящих к построению клики, для чего используется дополнительное множество, в которое помещаются вершины, бывшие уже использованными для увеличения полного подграфа.

Алгоритм оперирует тремя множествами вершин графа:

1. Множество **compsub** — множество, содержащее на каждом шаге рекурсии полный подграф для данного шага. Строится рекурсивно.

2. Множество **candidates** — множество вершин, которые могут увеличить compsub.

3. Множество **not** — множество вершин, которые уже использовались для расширения compsub на предыдущих шагах алгоритма.

Алгоритм является рекурсивной процедурой, применяемой к этим трем множествам. Сам алгоритм можно описать следующей последовательностью действий:

ПРОЦЕДУРА *extend* (*candidates, not*):

ПОКА *candidates* не пусто **И** *not* не содержит вершины, соединенной со всеми вершинами из *candidates*,

ВЫПОЛНЯТЬ:

1 Выбираем вершину *v* из *candidates* и добавляем её в *compsub*

2 Формируем *new_candidates* и *new_not*, удаляя из *candidates* и *not* вершины, не соединенные с *v*

3 **ЕСЛИ** *new_candidates* и *new_not* пусты

4 **ТО** *compsub* – клика

5 **ИНАЧЕ** рекурсивно вызываем *extend* (*new_candidates, new_not*)

6 Удаляем *v* из *compsub* и *candidates*, и помещаем в *not*

Найденные таким образом клики заданного размера могут быть использованы для описания социальных групп («семей») без внесения дополнительных изменений в структуру графов, что позволит сохранить их структурные особенности.

Использование случайного закона при определении соответствия между вершинами различных социальных графов лишает реалистичности синтезируемые массивы условно-реальных данных. Для максимального сохранения взаимосвязей между объектами в различных сервисах предложено использовать метод анализа структур социальных гра-

фов на основе дифференциации вершин и определения частичного изоморфизма [7].

Задача определения сходства структур рассматривается как выделение в сравниваемых графах G_M и G_S наибольшей общей части — графа $G_{max} = (U_{max}, E_{max})$. Для решения данной задачи в работах [8,9] предлагается перебрать все возможные подстановки вершин исследуемых графов, и в каждой из них определить число совместившихся ребер, образующих общую часть. Подстановка, образующая граф G_{max} с наибольшим числом рёбер $|E_{max}|$, именуется подстановкой сходства, а сформированная на основе нее общая часть является наибольшей.

С целью уменьшения вычислительной сложности разрабатываемого алгоритма предлагается исключить перебор всех возможных подстановок вершин при формировании наибольшей общей части путем введения начальной подстановки, которая определяет всю дальнейшую подстановку сходства за счет применения метода дифференциации вершин.

Под начальной подстановкой будем понимать пару вершин, для которых принимается условие взаимно однозначного соответствия (биективного отображения), т.е. для пары вершин $u_1(G_M) \in U_M$ и условно соответствующей ей $u_1(G_S) \in U_S$ верно $u_1(G_M) \in u_1(G_S)$.

В качестве начальной подстановки $(u_x(G_M), u_y(G_S))$ предлагается использовать вершины с максимальными степенными параметрами в обоих графах G_M и G_S:

$$u_1(G_M) = u_x^k(G_M) \in U_M$$
$$u_1(G_S) = u_y^n(G_S) \in U_S,$$

где k и n обозначают максимальные степени вершин $u_x(G_M)$ и $u_y(G_S)$ графов G_M и G_S соответственно.

После определения начальной подстановки выполняется процедура взаимозависимой дифференциации вершин в обоих графах G_M и G_S. В результате данного этапа происходит присвоение одинаковых кодов различия двум вершинам из различных графов, которые максимально соответствуют друг другу структурно. Таким образом, часть взаимосвязей пользователей из социального графа U_M будет доступна и в графе U_S. После определения соответствия вершин U_S вершинам U_M при необходимости происходит добавление ребер для сохранения ситуационной задачи Gt.

Формирование динамических компонентов массивов условно-реальных данных

ЦСП [10] представляет собой кортеж $CPN = <P, T, TM, I, O, M>$, где P — конечное множество позиций, T — конечное множество переходов, TM — множество временных моментов для срабатывания переходов, I — конечное множество входящих в переходы дуг, O — конечное множество выходящих из переходов дуг, M — множество меток в начальный момент времени.

Представленные статические и динамические компоненты массивов данных соответствуют объектам сетей Петри: динамические — изображаются метками (фишками, маркерами) внутри позиций и статические — им соответствуют вершины и дуги сети Петри.

Для создания вершин $(p_k \in P, t_k \in T)$ и дуг $(i_k \in I, o_k \in O)$ сети Петри используются принципы моделирования сложных сетей. Выбор моделей основывается на соответствии структурных параметров создаваемых графов статистическим распределениям, полученным в результате анализа реальных ИТ-сервисов.

Для отражения динамических свойств в сеть Петри введено понятие разметки сети, которая реализуется с помощью меток $m_k \in M$, размещаемых в позициях. Метка представляет собой объект, содержащий несколько параметров, каждый из которых может принимать дискретный набор значений. В соответствии с этим метки различаются по типам параметров (переменных). Чтобы отличать метки различных типов, их можно окрашивать в различные цвета. В задаче моделирования взаимодействия между пользователями ИТ-сервисов метки представляют собой коммуникационные события. Одним из ключевых параметров метки выступает тип коммуникационного события. Для ИТ-сервиса мобильной связи целесообразно использовать следующие типы событий [1]:

- *B* — отправка сообщений о подключениях к базовым станциям для передачи координат;
- *G* — получение *GPRS*-трафика;
- *C* — совершение звонка (разделяется на исходящие и входящие звонки);
- *S* — отправка СМС-сообщения (разделяется на исходящие и входящие СМС-сообщения).

При определении типов коммуникационных событий для ИТ-сервиса социальных сетей учитывается возможность пользователя вести публичную (тип – *public*) и приватную (тип – *private*) переписку. В зависимости от со-

держимого сообщения выделяются следующие типы событий:

- *text* — текстовое сообщение;
- *picture* — передача изображения;
- *video* — передача видео;
- *link* — ссылка на источник.

Для различных типов коммуникационных событий могут быть использованы необходимые дополнительные параметры, которые также будут содержаться в метке (например, продолжительность телефонного звонка). Таким образом, тип коммуникационного события является составным. Все параметры метки определяют ее составной цвет. Представление же самой метки $m_k \in M$, окрашенной цветом c*olour={private, text}*, имеет следующий вид: $m_k^{colour:\{private,\ text\}}$.

Сеть Петри представляет собой асинхронную систему, в которой метки перемещаются по позициям через переходы. Переход может сработать (т.е. переместить метку из входной позиции в выходную для данного перехода), если во всех входных позициях для данного перехода присутствует хотя бы одна метка и выполнено логическое выражение, ограничивающее переход (*спусковая функция*).

Дуги могут иметь пометки в виде выражений (переменных, констант или функций), определенных для множества цветов, и использоваться либо для «вычленения» компонентов сложного цвета меток при определении условия срабатывания перехода, либо для изменения цвета метки следующей позиции после срабатывания перехода. Данное свойство позволяет упростить процесс разделения коммуникационных событий на «семейные» (т.е. между членами социальной группы) и «все остальные» путем добавления нового параметра метки в зависимости от социального статуса получателя сообщения (тип – *family/others*) и создания различных условий срабатывания перехода между «членами семьи» и другими пользователями. Обозначение дуги $i_k \in I$, обеспечивающей срабатывание перехода меткой $m_k^{colour:\{family\}}$ с параметром *colour:{family}*, имеет следующий вид: $i_k^{colour:\{family\}}$.

Чаще всего таблицы взаимодействия пользователей ИТ-сервисов наполняются записями в порядке их совершения, т.е. упорядочены по времени начала события. Время представляется в формате UTC, минимальным шагом изменения состояния является 1 секунда. Для анализа систем реального времени введен временной механизм, реализованный с помощью глобальных часов и так называемых *штампов*, которые несут метки. Временной штамп метки назначается при ее инициализации в начальной разметке и наращивается выражениями на переходах или дугах. В результате метка становится доступной для перехода, если ее штамп оказался меньше значения счетчика глобальных часов. Обозначение временного штампа метки $m_k \in M$, активирующей переход во временной момент *time:{z}*, $z \in TM$, принимает следующий вид: $m_k^{time:\{z\}}$. Часы наращивают свое значение, если на данный момент времени ни один переход сети не разрешен.

Конкретизация метки, находящейся в данной позиции, определяется инициализирующим выражением начальной разметки. Для выполнения начальной разметки требуется определить некоторые значения:

- общее количество меток $|M|=K$;
- набор параметров, определяющих цвет каждой метки $m_k^{colour:\{c1,\ c2,\ \ldots,\ cP\}}$, где *{c1, c2, …, cP}* — типовые и дополнительные параметры коммуникационного события *k, k=1, …, K*;
- временные штампы меток $m_k^{time:\{z\}}$, где $z \in TM$.

Для формирования начальной разметки ЦСП, описывающей взаимодействие пользователей ИТ-сервиса мобильных сетей, целесообразно использовать статистические характеристики биллинговой информации [1]:

- $K_0 = \langle F_{time}, F_{dur}, F_n, F_a \rangle$ — не связанные с адресацией соединения, описываемые функциями распределения: F_{time} — времени суток, F_{dur} — длительности события, F_n — количества соединений, F_a — вероятности генерации типа события;
- $F_1 \langle F_l, F_t \rangle$ — связанные с адресацией соединения: F_l — функция распределения выбора получателей соединения, а F_t — функция распределения промежутков времени между началами инициализации двух последовательных соединений.

В результате инициализации начальной разметки сети происходит распределение меток с заданными параметрами по определенным позициям в соответствии со статистическими распределениями различных характеристик рассматриваемых ИТ-сервисов. На основе выбранного временного интервала для описания взаимодействия пользователей ИТ-сервисов запускается механизм глобальных часов. После чего происходит последовательное перемещение меток между

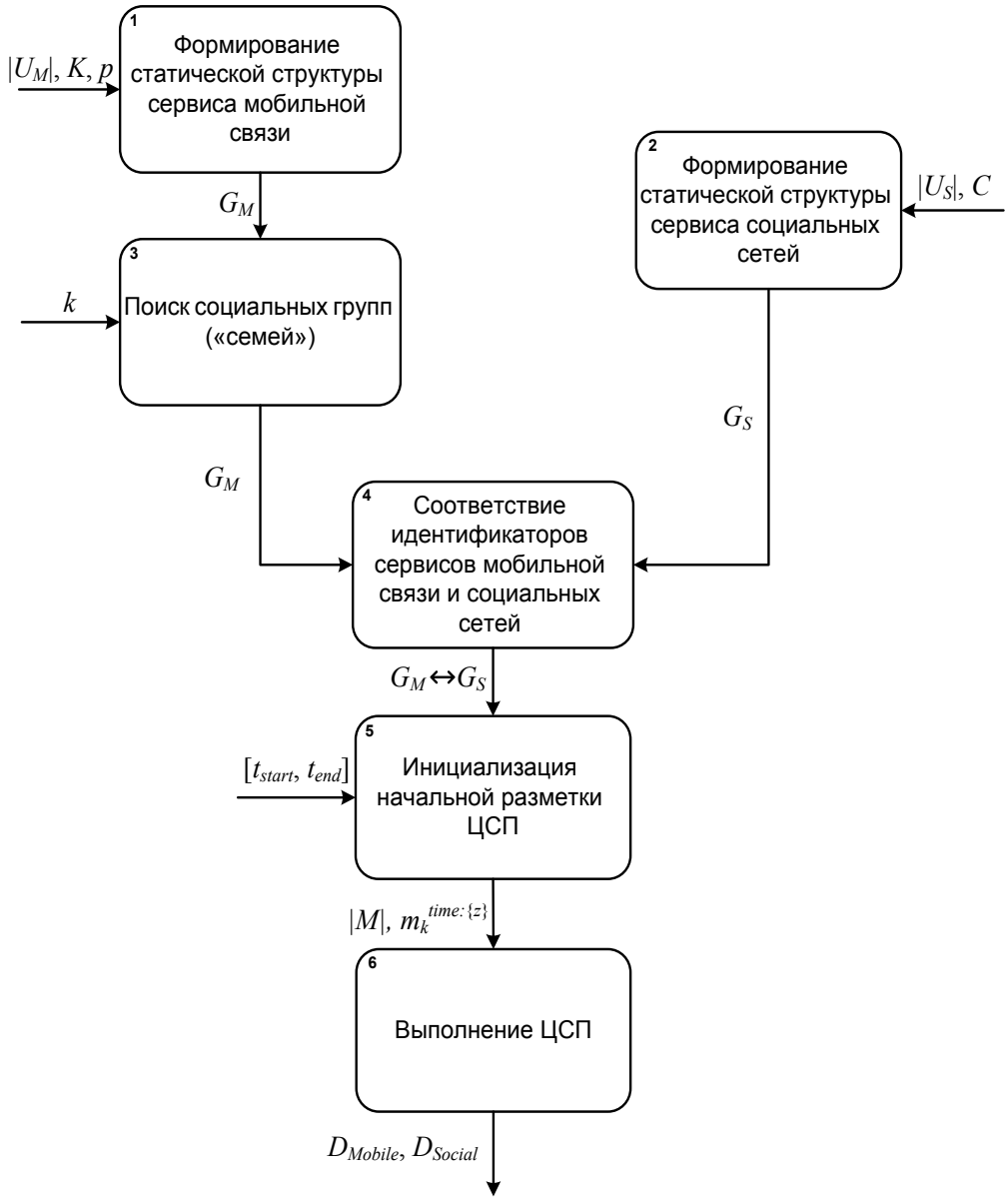


Рис. 1. Структурная схема комплексного метода синтеза массивов данных

позициями, что позволяет создавать строки d_{Mobile} и d_{Social} в массивах D_{Mobile} и D_{Social}.

Таким образом, комплексный метод синтеза массивов условно-реальных данных представляет собой последовательность математических подходов, имитирующих различные параметры процесса взаимодействия пользователей на основе статистических распределений требуемых показателей в реальных ИТ-сервисах. Структурная схема метода синтеза представлена на рис. 1.

1. Для формирования статической структуры социального графа сервиса мобильной связи G_M в соответствии с моделью Ваттца-Строгатца определяются три основных параметра:

– численность пользователей имитируемого сервиса $|U_M|$,

– среднее количество социальных связей пользователей, описываемое через степень вершин K создаваемой регулярной решетки,

– вероятность перераспределения каждого ребра на случайную вершину p.

Модификация значений K и p позволяет менять такие структурные свойства создаваемого графа как минимальная длина пути между вершинами и коэффициент кластеризации. Для имитации сети вербального обще-

ния людей в соответствии со статистически определенными структурными параметрами в работах [4, 5] оптимальными значениями выбраны K=6 и p=0,1.

2. При создании статической структуры сервиса социальных сетей G_S в соответствии с моделью Барабаши-Альберт необходимо определить следующие параметры:

– численность пользователей имитируемого сервиса $|U_S|$,

– количество ребер C, с которыми новая вершина добавляется в структуру графа.

Параметр C определяет важное свойство создаваемого графа – степень кластеризации вершин. Для моделирования структуры социальных сетей в соответствии со статистическими параметрами в [5] достаточным значением является C=3.

3. Поскольку вовлеченность людей в сервис мобильной связи значительно выше, чем в социальные сети, именно в графе G_M производится поиск клик заданного размера в соответствии с алгоритмом Брона-Кербоша. Единственным параметром, задаваемым для алгоритма, является размер клики k. Для соответствия распределению количества одиноких людей и реальных семей с различным составом детей (0-3), последовательно производится поиск клик размерностью k=3, k=4, k=5 в требуемом соотношении.

4. Для имитации одновременного общения пользователей посредством различных ИТ-сервисов применяются методы анализа социальных графов G_M и G_S, направленные на установление соответствия персональных идентификаторов сервисов мобильной связи и социальных сетей *(AbonentIMSI, AbonentIMEI, AbonentPhone)* ↔ (AbonentUID, AbonentIP, AbonentLogin)

5. Для инициализации начальной разметки ЦСП необходимо определить временной интервал $[t_{start}, t_{end}]$, в течение которого происходит моделирование взаимодействия между пользователями. На основании данных параметров произойдет автоматическое формирование общего количества меток $|M|$ и временных штампов меток $m_k^{time:\{z\}}$, а также распределение их по позициям $p_k \in P$ в соответствии со статистическими распределениями активности пользователей для указанного временного промежутка.

6. При выполнении циклов ЦСП имитируется информационный обмен между заданным количеством пользователей $|U_M|$ и $|U_S|$ в определенный временной период $[t_{start}, t_{end}]$. Результатом выполнения всех циклов ЦСП являются синтезированные массивы D_{Mobile} и D_{Social} требуемого формата.

Анализ синтезированных массивов условно-реальных данных с применением ИАСБ на базе ПО Lampyre

Синтезированные массивы DMobile и DSocial имеют необходимый формат для интеграции в учебные ИАСБ компьютерного полигона. Обычно это файлы форматов *.xml и *.csv.

В составе учебного полигона включены различные аппаратно-программные комплексы для проведения поисково-аналитической работы, одним из которых является ПО Lampyre. Данная система позволяет обрабатывать большие массивы данных, строить графики взаимосвязей и накладывать их на карту и временной масштаб.

На компьютерном полигоне для анализа сетевого взаимодействия используются разрабатываемые студентами в процессе обучения аналитические методики, которые представляют собой программный код на языке Python, интегрируемый в ПО Lampyre. Данные методики и различные варианты визуализации графа связей, используемые в ПО Lampyre, применяются к синтезированному массиву данных. Они позволяют решить сквозную аналитическую задачу по поиску в массиве данных источника атакующего воздействия, ставшего причиной инцидента информационной безопасности, на основе анализа сетевого взаимодействия. На рис. 2 представлен интерфейс ПО Lampyre, визуализирующий граф связей условного пользователя при решении учебной аналитической задачи.

Заключение

В статье рассмотрен комплексный имитационно-статистический метод синтеза массивов условно-реальных данных позволяющий формировать статические и динамические компонентов. Предложенный композиционный метод, объединяющий несколько моделей построения сложных сетей для формирования статических структур моделируемых ИТ-сервисов, используются в качестве основы для создания коммуникационных событий. В результате применения алгоритмов ЦСП в процессе синтеза массивов данных имеется возможность создания записей о взаимодействии пользователей, наполнение которых зависит от структурных, событий-

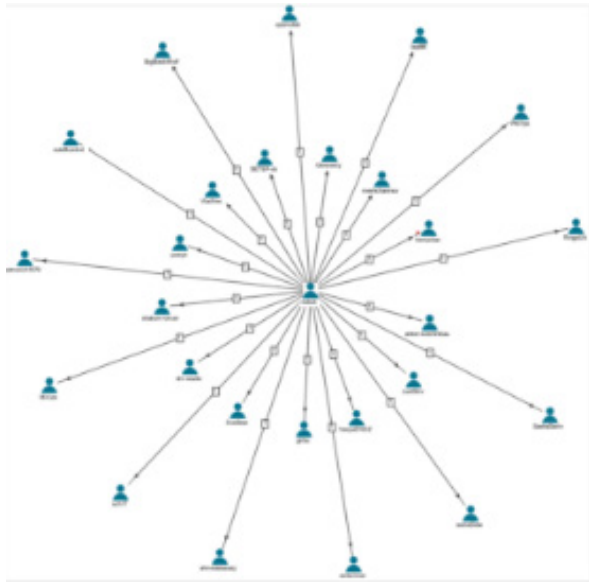


Рис. 2. Интерфейс ПО Lampyre

ных, временных и социальных параметров моделируемых ИТ-сервисов. Сгенерированный ситуационный массив условно-реальных данных применяется при отработке практических заданий по разработке поисково-аналитических методик на учебном компьютерном полигоне по расследованию инцидентов информационной безопасности.

Литература

1. Семенищев И.А., Синадский А.Н., Синадский Н.И., Сушков П.В. Синтез массивов биллинговой информации на основе статистико-событийной модели взаимодействия абонентов сетей сотовой связи. // Вестник УрФО. Безопасность в информационной сфере. — 2018. — № 1 (27). — С. 47–56.

2. Erdos, P., and A. Renyi, On Random Graphs. Publicationes Mathematicae (Debrecen), volume 6, 1959, pp. 290–297.

3. Watts, D. J., Small Worlds: The Dynamics of Networks between Order and Randomness (Princeton University, Princeton, NJ), 1999.

4. R. Albert, A-L. Barabasi. Statistical mechanics of complex networks. Reviews of modern physics, volume 74, January 2002.

5. Проект Edyo.ru. URL: http://edyo.ru (дата обращения 25.03.2020).

6. C. Bron and J. Kerbosch, Algorithm 457: Finding All Cliques of an Undirected Graph, Proceedings of the ACM, 1973, p. 575–577.

7. Синадский Н.И., Сушков П.В. Модификация методов анализа социальных графов на основе применения атрибутивных компонентов учетных записей для идентификации сообществ пользователей социальных сетей. — Вестник УрФО. Безопасность в информационной сфере. — 2017. — № 2 (24). — С. 32–40.

8. Погребной Ан.В., Погребной В.К. Метод дифференциации вершин графа и решение проблемы изоморфизма // Известия Томского политехнического университета. — 2015. — Т. 326. — № 6. — С. 34–45.

9. Погребной А.В. Метод определения сходства структур графов на основе выделения частичного изоморфизма в задачах геоинформатики // Известия Томского политехнического университета, 2015. — Т. 326. — № 11. — С. 56–66.

10. Питерсон Дж. Теория сетей Петри и моделирование систем: пер. с англ. — М.: Мир, 1984. — 264 с.

References

1. Semenishchev I.A., Sinadskiy A.N., Sinadskiy N.I., Sushkov P.V. Sintez massivov billingovoy informatsii na osnove statistiko-sobytiynoy modeli vzaimodeystviya abonentov setey sotovoy svyazi. — Vestnik UrFO. Bezopasnost' v informatsionnoy sfere. — 2018. № 1 (27). — S. 47-56.

5. Project Edyo.ru. URL: http://edyo.ru (data obrashcheniya 25.03.2020).

7. Sinadskiy N.I., Sushkov P.V. Modifikatsiya metodov analiza sotsial'nykh grafov na osnove primeneniya atributivnykh komponentov uchetnykh zapisey dlya identifikatsii soobshchestv pol'zovateley sotsial'nykh setey. — Vestnik UrFO. Bezopasnost' v informatsionnoy sfere. — 2017. № 2 (24). — S. 32-40.

8. Pogrebnoy An.V., Pogrebnoy V.K. Metod differentsiatsii vershin grafa i reshenie problemy izomorfizma // Izvestiya Tomskogo politekhniche-skogo universiteta. — 2015. — T. 326. — № 6. — S. 34–45.

9. Pogrebnoy A.V. Metod opredeleniya skhodstva struktur grafov na osnove vydeleniya chastichnogo izomorfizma v zadachakh geoinformatiki // Izvestiya Tomskogo politekhnicheskogo universiteta, 2015. — T. 326. — № 11. — S. 56-66.

10. Piterson Dzh. Teoriya setey Petri i modelirovanie sistem: Per. s angl. — M.: Mir, 1984. — 264 s.

ГАЙДАМАКИН Николай Александрович, доктор технических наук, профессор, профессор учебно-научного центра «Информационная безопасность», Институт радиоэлектроники и информационных технологий - РТФ Уральский федеральный университет им. первого Президента России Б.Н. Ельцина. 620002; г. Екатеринбург, ул. Мира, 19. E-mail: n.a.gaidamakin@urfu.ru

СИНАДСКИЙ Николай Игоревич, кандидат технических наук, доцент, доцент учебно-научного центра «Информационная безопасность», Институт радиоэлектроники и информационных технологий - РТФ Уральский федеральный университет им. первого Президента России Б.Н. Ельцина. 620002, г. Екатеринбург, ул. Мира, 19. E-mail: n.i.sinadsky@urfu.ru

СУШКОВ Павел Владимирович, аспирант ИЕНМ, Институт радиоэлектроники и информационных технологий - РТФ Уральский федеральный университет им. первого Президента России Б.Н. Ельцина. 620002, г. Екатеринбург, ул. Мира, 19. E-mail: pavelsu@e1.ru

GAIDAMAKIN Nikolay, doctor of engineering, Professor, Educational and Scientific Center "Information Security", Institute of Radio electronics and Information Technologies, Ural Federal University named after first President of Russia B.N. Yeltsin; 620002, Yekaterinburg, Mira str., 19. E-mail: n.a.gaidamakin@urfu.ru

SINADSKY Nikolay, candidate of technical sciences, associate Professor, Educational and Scientific Center "Information Security", Institute of Radio electronics and Information Technologies, Ural Federal University named after first President of Russia B.N. Yeltsin; 620002, Yekaterinburg, Mira str., 19. E-mail: n.i.sinadsky@urfu.ru

SUSHKOV Pavel, post-graduate student of Institute of Natural Sciences and Mathematics, Ural Federal University named after first President of Russia B.N. Yeltsin; 620002, Yekaterinburg, Mira str., 19. E-mail: pavelsu@e1.ru

УДК 004.896 + 004.056.5 Вестник УрФО № 1(35) / 2020, с. 24–34

Рагозин А. Н.

DOI: 10.14529/secur200103

ПРИМЕНЕНИЕ ЦИФРОВОЙ ОБРАБОТКИ СИГНАЛОВ И НЕЙРОННОЙ СЕТИ ПРИ ФОРМИРОВАНИИ ПРОГНОЗА ВРЕМЕННЫХ РЯДОВ ДАННЫХ ДЛЯ ЦЕЛЕЙ ОБНАРУЖЕНИЯ АНОМАЛИЙ ПРИ АВТОМАТИЗИРОВАННОМ УПРАВЛЕНИИ ТЕХНОЛОГИЧЕСКИМИ ПРОЦЕССАМИ

С целью обнаружения аномалий и повышения качества прогнозирования динамических потоков данных, наблюдаемых с сенсоров в автоматизированных системах управления технологическими процессами (АСУ ТП), предлагается применять прогнозирующий модуль, состоящий из последовательно соединенных блока цифровой обработки сигналов (ЦОС) и прогнозирующего блока с использованием нейронной сети (НС) (прогнозирующего автоэнкодера (автокодировщика), predictive Autoencoder (PAE)). В проведенном исследовании показано, что блок предварительной ЦОС входного прогнозируемого сигнала, состоящий из параллельного набора (гребенки) цифровых фильтров нижних частот с конечными импульсными характеристиками (КИХ-ФНЧ), приводит к неравновесному учету корреляционных связей временных отсчетов входного сигнала и повышению точности конечного результата прогнозирования. Предложенный и рассмотренный в работе прогнозирующий автоэнкодер (PAE), кроме восстановления на выходе PAE входного сигнала, или части входного сигнала, также формирует на выходе прогнозируемые отсчеты входного сигнала на заданное количество временных шагов «вперед», что повышает точность результата прогнозирования. Уменьшение ошибки прогноза происходит за счет наложения ограничений при формировании прогноза, то есть дополнительного требования восстановления на выходе НС входных выборок отсчетов – «стабилизаторов». Введение «стабилизаторов» повышает точность результата прогнозирования. При возникновении анома-

лий в работе АСУ ТП будут происходить структурные изменения в сигнале ошибки формируемого прогноза, в результате анализа этих структурных изменений ошибки прогноза, собственно происходит детектирование (обнаружение) аномалий в наблюдаемых процессах АСУ ТП. В данном случае рассматривается режим обучения «частично с учителем». Исходные данные при таком подходе представляют только класс «нормальных данных», при этом, при обучении «частично с учителем» не требуется информация об аномальном классе целевых данных

Ключевые слова: *цифровая фильтрация, нейронная сеть, прогнозирование, прогнозирующий автоэнкодер, автокодировщик, вертикальный сигнал, автоматизированная система управления технологическими процессами, поток данных.*

Ragozin A. N.

THE USE OF DIGITAL SIGNAL PROCESSING AND A NEURAL NETWORK WHEN GENERATING A FORECAST OF TIME SERIES OF DATA FOR THE PURPOSE OF DETECTING ANOMALIES IN THE IN THE AUTOMATED CONTROL OF TECHNOLOGICAL PROCESSES

In order to detect anomalies and improve the quality of forecasting dynamic data flows observed from sensors in Industrial Control System (ACS)., it is proposed to use a predictive module consisting of a series-connected digital signal processing unit (DSP) and a predictive unit using a neural network (predictive autoencoder (Auto Encoder), predictive Autoencoder (PAE)). The study showed that the preliminary DSP block of the predicted input signal, consisting of a parallel set (comb) of digital low-pass filters with finite impulse responses (FIR-LPF), leads to a non-equilibrium account of the correlation relationships of the time samples of the input signal and to increase the accuracy of the final prediction result. The predicted autoencoder (PAE) proposed and considered in the work, in addition to restoring the input signal or part of the input signal at the PAE output, also generates the predicted samples of the input signal for the specified number of «forward» time steps at the output, which increases the accuracy of the prediction result. The reduction of the forecast error occurs due to the imposition of restrictions in the formation of the forecast, that is, an additional requirement to restore the input samples of the samples – «stabilizers» at the NS output. The introduction of «stabilizers» increases the accuracy of the prediction result. When anomalies occur in the ACS operation, structural changes will oc-

cur in the error signal of the generated forecast, as a result of the analysis of these structural changes in the forecast error, the anomalies are actually detected in the observed ACS processes. In this case, the mode of study is «partly with the teacher». The initial data in this approach represent only the class of «normal data», while learning «partly with the teacher» doesn't require information about the anomalous class of target data.

Keywords: *digital filtering, neural network, forecasting, predictive autoencoder (Auto Encoder), vertical signal, Industrial Control System, data stream.*

Введение

Обнаружение аномалий в процессах (в динамических потоках данных) во многом определяет эффективность управления информационной безопасностью в автоматизированных системах управления технологическими процессами (АСУ ТП). Методы обнаружения аномалий применяются для решения задач обнаружения атак как на информационном уровне, так и на кибернетическом уровне в АСУ ТП.

Атаки вызывают аномалии (то есть, неожидаемое изменение) в поведении наблюдаемых процессов (в динамике наблюдаемых временных рядов данных) при работе АСУ ТП. Методы обнаружения аномалий в процессах АСУ ТП относят к поведенческим методам [1,2]. Поведенческие методы основаны на моделях «нормального» функционирования АСУ ТП. При этом задача обнаружения аномалий состоит в обнаружении расхождений между текущим (наблюдаемым) процессом работы АСУ ТП и процессом работы, который является эталонным для АСУ ТП (то есть, для АСУ ТП, работающей в штатном режиме). Любое несоответствие наблюдаемого процесса и эталонного процесса АСУ ТП рассматривается как аномалия (или вторжение).

Для обнаружения аномалий (или, вторжений) в АСУ ТП применяются нейронные сети, и с их использованием процедура прогнозирования наблюдаемых динамических потоков данных [3–7]. При построении прогноза нейронная сеть обучается при нормальной, штатной (без влияния дестабилизирующих воздействий) работе АСУ ТП. Детектирование (обнаружение) аномалий происходит в результате сравнения наблюдаемого временного ряда (динамического потока данных) АСУ ТП, с прогнозом этого временного ряда (прогнозом динамического потока данных), формируемого нейронной сетью, раннее обученной при нормальной работе АСУ ТП (работе АСУ ТП в штатном режиме, без влияния дестабилизирующих воздействий). Результат сравнения (разности) наблюдаемого и прогнозируемого временного ряда (наблюдаемого процесса) АСУ ТП определяется как сигнал (временной ряд) ошибки прогноза.

При возникновении аномалий в работе АСУ ТП будут происходить структурные изменения в сигнале ошибки формируемого прогноза, в результате анализа этих структурных изменений ошибки прогноза, собственно происходит детектирование (обнаружение) аномалий в наблюдаемых процессах АСУ ТП.

В данном случае рассматривается режим обучения «частично с учителем». Исходные данные при таком подходе представляют только класс «нормальных данных», при этом, при обучении «частично с учителем» не требуется информация об аномальном классе целевых данных

Очевидно, что точность настройки формируемого прогноза наблюдаемого процесса АСУ ТП, работающей в штатном режиме играет важную роль, так как в данном случае формируемый прогноз наблюдаемого процесса АСУ ТП является эталонной моделью процесса АСУ ТП, работающей в штатном режиме [8].

1. Актуальность и прикладная значимость формирования прогноза при обнаружении аномалий в процессах АСУ ТП.

Процессы, протекающие в АСУ ТП, являются сложными многокомпонентными процессами. Прогноз сложных многокомпонентных временных рядов данных необходимо проводить в различных временных масштабах (краткосрочный, среднесрочный, долгосрочный), то есть необходимо формировать многокомпонентный прогноз временных рядов. Этапы построения многокомпонентного прогноза временных рядов данных рассмотрены в работах [9, 10]. При этом наблюдаемый сложный многокомпонентный процесс АСУ ТП раскладывается с использованием технологии цифровой обработки сигналов (ЦОС) на отдельные базовые компоненты, далее каждая базовая компонента прогнозируется на свой интервал прогнозирования с ис-

пользованием технологии нейронных сетей, в итоге по совокупности полученных прогнозов с различными интервалами прогнозирования формируется завершенный многокомпонентный прогноз наблюдаемого процесса работающей АСУ ТП. Методы прогнозирования, также рассмотрены в работах [11-15].

Формирование многокомпонентного прогноза позволяет детектировать (обнаруживать) аномалии в наблюдаемом множестве всех временных рядов, отражающих работу АСУ ТП в результате покомпонентного сравнения каждой отдельной компоненты многокомпонентного временного ряда с каждым прогнозом соответствующей отдельной компоненты этого временного ряда, что позволит обнаруживать во множестве наблюдаемых временных рядов АСУ ТП аномалии раздельно по их различным динамическим характеристикам. Подобный подход существенно повышает эффективность управления информационной безопасностью в АСУ ТП.

Необходимо отметить, что для целей обнаружения аномалий в работе АСУ ТП к формируемому прогнозу, наблюдаемых временных рядов АСУ ТП предъявляются высокие требования по качеству формируемого прогноза.

В работе исследуется структура прогнозирующего модуля и влияние ее параметров на качество прогноза наблюдаемых сигналов АСУ ТП.

2. Формирование структуры прогнозирующего модуля.

В работе рассматривается и исследуется прогнозирующий модуль, состоящий из последовательно соединенных блока ЦОС и прогнозирующего блока с использованием нейронной сети (прогнозирующего автоэнкодера (автокодировщика), predictive Autoencoder (PAE)).

Исследуется влияние параметров блока ЦОС на качество прогнозирования входного сигнала, также исследуется влияние на качество формируемого прогноза, особенности архитектуры прогнозирующего автоэнкодера (PAE), налагающей дополнительные ограничения на результат формируемого прогноза в виде требований дополнительного восстановления на выходе PAE входного сигнала, или части входного сигнала.

Для формирования прогноза наблюдаемого временного ряда (или, отдельной компоненты наблюдаемого многокомпонентного временного ряда) предлагается к рассмотрению прогнозирующий модуль, состоящий из последовательно соединенных блока ЦОС и прогнозирующего блока с использованием нейронной сети (рис. 1).

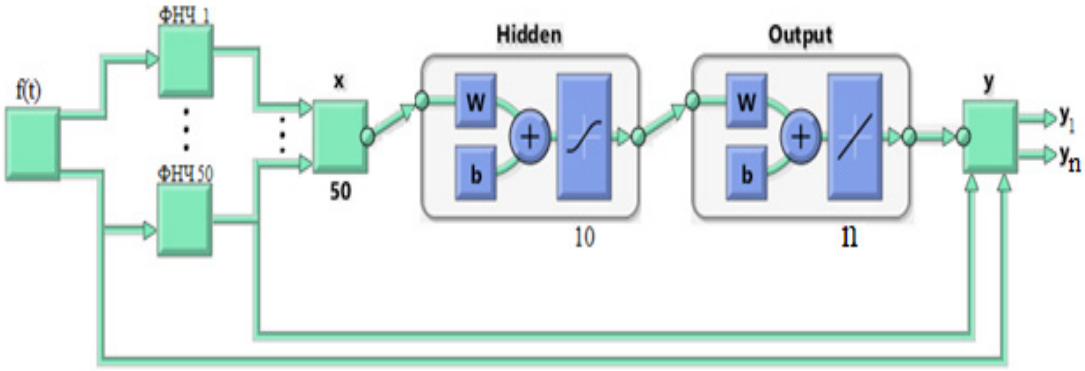


Рис. 1. Структурная схема прогнозирующего модуля: блок ЦОС и прогнозирующий блок с использованием нейронной сети

В прогнозирующем модуле (рис. 1) блок ЦОС состоит из гребенки цифровых фильтров нижних частот (ФНЧ). При этом на выходе гребенки ФНЧ формируется набор отфильтрованных компонентов входного сигнала. Набор полученных отфильтрованных компонентов входного сигнала с выхода гребенки ФНЧ назовем «вертикальным сигналом». То есть, «вертикальный сигнал», это многоканальный сигнал с выхода гребенки ФНЧ, в данном примере (рис. 1), состоящей из 50 КИХ-ФНЧ (ФНЧ с конечной импульсной характеристикой) с последовательно уменьшающимися частотами среза их частотных характеристик.

При таком подходе происходит последовательная фильтрация шумов входного сигнала с использованием параллельного набо-

ра (гребенки) цифровых КИХ-ФНЧ. В данном случае, более далекие (с большим запаздыванием по времени) от текущего момента времени отсчеты входного сигнала подвергаются более глубокому сглаживанию, чем отсчеты входного сигнала, более близкие к текущему моменту времени, что в свою очередь приводит к неравновесному учету корреляционных связей временных отсчетов входного сигнала и текущего отсчета результата прогноза на выходе прогнозирующего модуля (рис. 1). Неравновесный учет корреляционных связей временных отсчетов входного сигнала в формируемом прогнозе приводит к повышению точности результата прогнозирования.

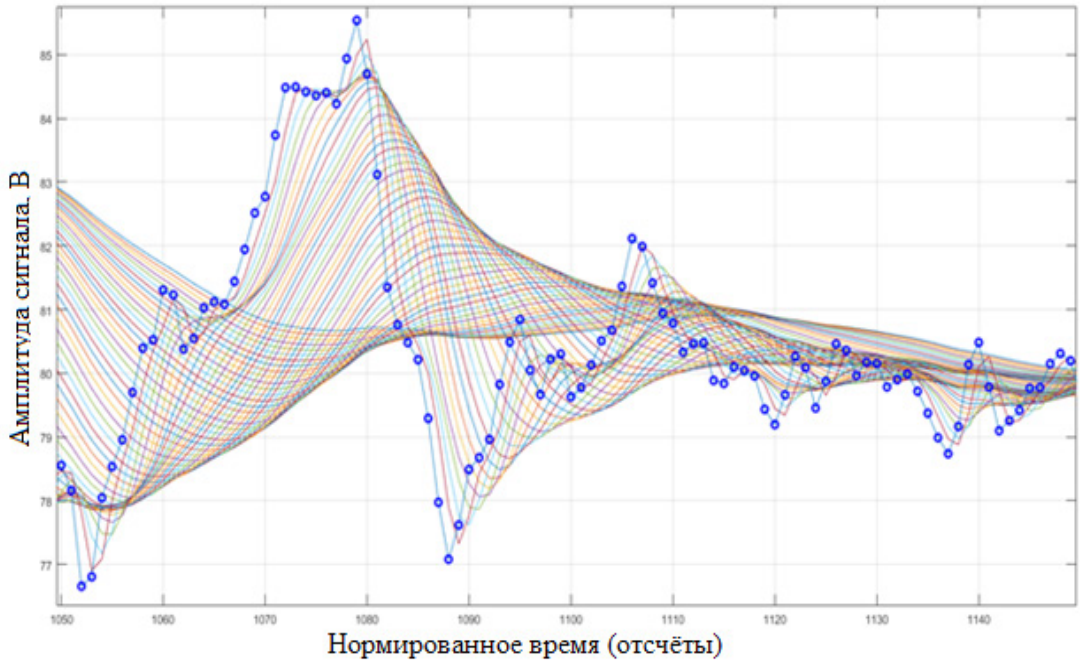


Рис. 2. Входной сигнал (синие точки) и отфильтрованные компоненты с выхода блока ЦОС

В более ранних исследованиях в ходе проведенных экспериментов установлено, что использование полученных дополнительных признаков из сформированного описанным способом «вертикального сигнала» повышает точность обнаружения кибератак на АСУ ТП [16].

На рис. 2, в качестве примера изображен входной сигнал (синие точки) и «вертикальный сигнал» – набор полученных отфильтрованных 50 компонентов входного сигнала с выхода гребенки 50 КИХ-ФНЧ с последовательно уменьшающимися частотами среза их частотных характеристик. Необходимо отметить, что с выхода блока ЦОС далее, на вход прогнозирующего блока с использованием нейронной сети подается последовательный набор («вертикальный сигнал» с количеством отсчетов равным 50), состоящий из последних временных отсчетов каждой из 50 полученной компоненты с выхода гребенка из 50 КИХ-ФНЧ.

В прогнозирующем модуле (рис. 1) прогнозирующий блок с использованием нейронной сети, следующий за блоком ЦОС состоит из двухслойной нейронной сети (в данном примере, скрытый слой – 10 нейронов, тангенциальная функция активации, выходной слой – линейная функция активации).

Необходимо отметить, что в прогнозируемом блоке может использоваться другая архитектура НС.

На вход прогнозирующего блока подаются отсчеты «вертикального сигнала» с выхода блока ЦОС, при этом, целевыми отсчетами (подаваемыми на выход НС) при обучении прогнозирующего модуля являются целевые (прогнозируемые) отсчеты входного сигнала (впоследствии, итоговые результаты прогноза), также, подаются дополнительно целевые отсчеты, выполняющие роль «стабилизаторов» – выборка заданной длины из отсчетов «вертикального сигнала» со входа НС (то есть, с выхода блока ЦОС), плюс выборка заданной длины из отсчетов исходного входного сигнала (то есть, со входа блока ЦОС).

Введение «стабилизаторов» наделяет прогнозирующий блок (рис. 1) свойствами автоэнкодера (автокодировщика) (Autoencoder (AE)) [17–20], что сужает пространство выбора при формировании прогноза НС за счет наложения ограничений при формировании прогноза. Ограничения требуют, кроме восстановления на выходе НС прогнозируемых отсчетов выборок сигнала (то есть, результата прогноза), также, дополнительно восстановление входных выборок отсчетов – «стабилизаторов». Введение «стабилизаторов» повышает точность результата прогнозирования.

В данном случае (при наличии дополнительных целевых отсчетов – «стабилизаторов») для прогнозирующего блока с использованием НС можно предложить название «прогнозирующего автоэнкодера» (predictive Autoencoder (PAE)).

Предложенная в работе структура PAE обобщает понятие классического автоэнкодера (Autoencoder (AE)), наделенного только свойством восстановления входного сигнала.

3. Исследование структуры и параметров блока ЦОС прогнозирующего модуля на точность формируемого прогноза.

Рассмотрим влияние блока ЦОС на точность формируемого прогноза.

На рис. 3 изображена структурная схема прогнозирующего модуля с упрощенной структурой: отсутствует блок ЦОС, на вход НС подаются непосредственно отсчеты входного сигнала (прогнозируемого сигнала), также отсутствует использование выборок стабилизаторов. То есть, в данном случае прогноз осуществляется непосредственно двухслойной НС, на вход которой подаются 50 последовательных отсчета входного (прогнозируемого) сигнала.

На рис. 4 изображена структурная схема прогнозирующего модуля с добавленным блоком ЦОС при этом, отсутствует использование выборок стабилизаторов. В данном случае прогноз входного сигнала осуществляется двухслойной НС, на вход которой подаются 50 последовательных отсчетов «вертикального сигнала» с выхода блока ЦОС из 50 КИХ – ФНЧ.

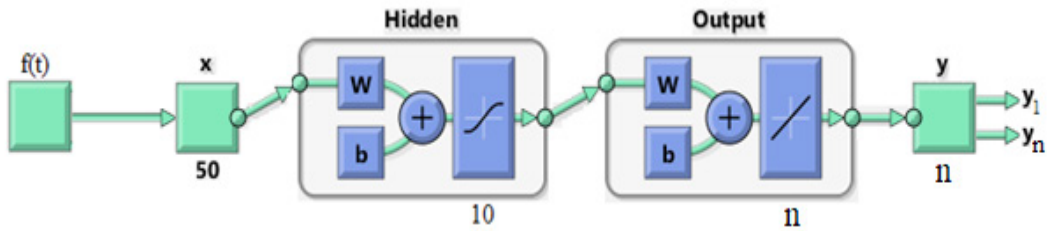


Рис. 3. Структурная схема прогнозирующего модуля: отсутствует блок ЦОС и отсутствует отсчеты – «стабилизаторы» в прогнозирующем блоке с использованием нейронной сети

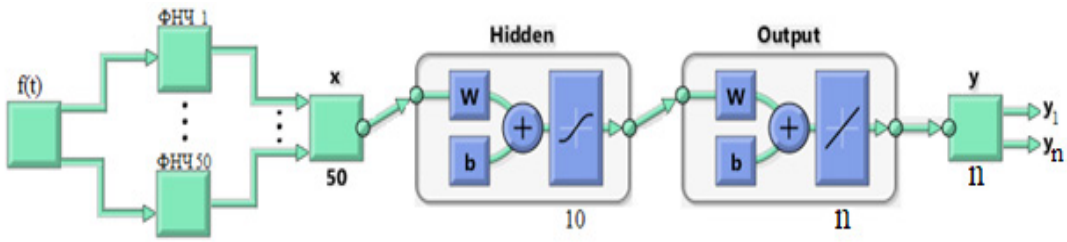


Рис. 4. Структурная схема прогнозирующего модуля: добавлен блок ЦОС, но отсутствует отсчеты – «стабилизаторы» в прогнозирующем блоке с использованием нейронной сети

Структура PAE, кроме восстановления входного сигнала реализует, также дополнительно его прогнозирование за пределы восстанавливаемого входного сигнала. Наделение архитектуры AE классического автоэнкодера также, свойствами «предиктора» (то есть, перевод архитектуры AE в архитектуру PAE), дает возможность существенно расширить область решаемых им задач, то есть задач, решаемых с использованием архитектуры PAE.

На рис. 5 приведены результаты формирования прогноза, исследуемого технического входного сигнала на один шаг и два временных шага вперед с использованием прогнозирующего модуля, представленного на рис. 3. На рис. 5, синяя линия – целевая линия, то есть линия сигнала, с которой значения прогноза должны совпадать (учтено смещение по времени для прогнозируемых и целевых отсчетов сигнала). Пурпурная линия –

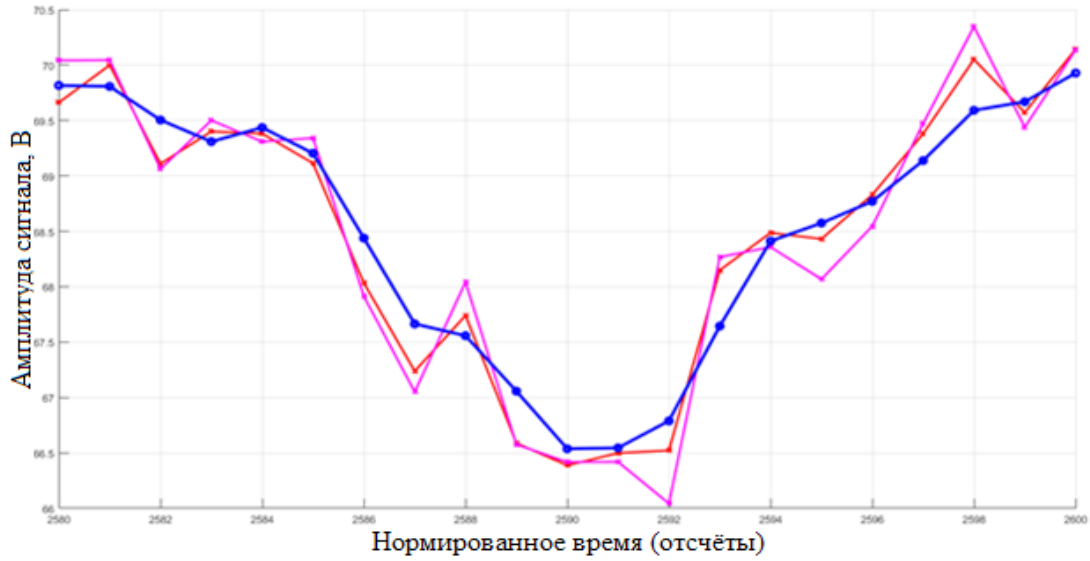


Рис. 5. Сформированный прогноз на один и два шага по времени вперед, отсутствует блок ЦОС

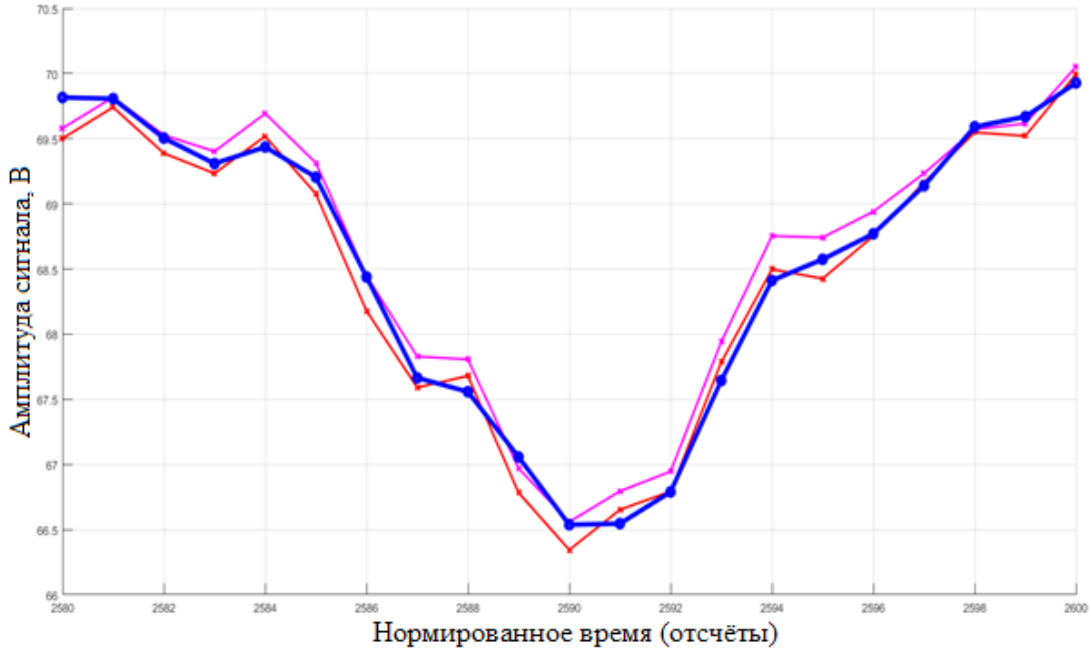


Рис. 6. Сформированный прогноз на один и два шага по времени вперед при наличии блока ЦОС

прогноз на один шаг по времени вперед, красная линия – прогноз на два шага по времени вперед.

На рис. 6 приведены результаты формирования прогноза входного сигнала на один шаг и два временных шага вперед с использованием прогнозирующего модуля, представленного на рис. 4, то есть при наличии блока ЦОС.

Сравнительный анализ результатов, представленных на рис. 5 и 6, свидетельствует о более высокой точности результата прогнозирования с использованием блока ЦОС в прогнозирующем модуле. Для данного примера среднеквадратичное отклонение (СКО) результата прогноза на два шага вперед, представленного на рис. 5 имеет величину 0,2101, а для результата прогноза, представленного на рис. 6 имеет величину СКО=0,1917, соответственно. На рис. 5, 6 приведены результаты подгонки прогнозирующего модуля (отображено в диапазоне от 2580 до 2600 нормированных временных отсчётов). Обучение НС методом Левенберга-Марквардa производилось на выборке сигнала длиной 5000 отсчетов, при этом распределение входных данных: подгонка – 70%, контроль – 15%, тестирование – 15%.

4. Исследование влияния структуры прогнозирующего блока с использованием нейронной сети на точность формируемого прогноза.

Рассмотрим влияние использования целевых отсчетов «стабилизаторов» на точность формируемого прогноза. В этом, случае формирование прогноза входного сигнала производится при использовании «полной» архитектуры прогнозирующего модуля (рис. 1).

На рис. 7 приведены результаты формирования прогноза входного сигнала (результаты тестирования, отображено в диапазоне от 2800 до 2870 нормированных временных отсчётов за пределами интервала подгонки) на один, два и три временных шага вперед с использованием прогнозирующего модуля, представленного на рис. 1, то есть при наличии блока ЦОС и использовании отсчетов «стабилизаторов». Для данного примера используется шесть отсчетов «стабилизаторов»: три последовательных отсчета с выхода блока ЦОС, то есть со входа НС и три последовательных отсчета входного сигнала, то есть со входа блока ЦОС.

На рис. 7, линии прогноза изображены с опережением по времени прогнозируемого сигнала, соответственно шагу прогноза по времени вперед.

Необходимо отметить, что в данном случае прогнозирующий блок (рис. 1) имеет архитектуру, предложенного в данной работе «прогнозирующего автоэнкодера PAE».

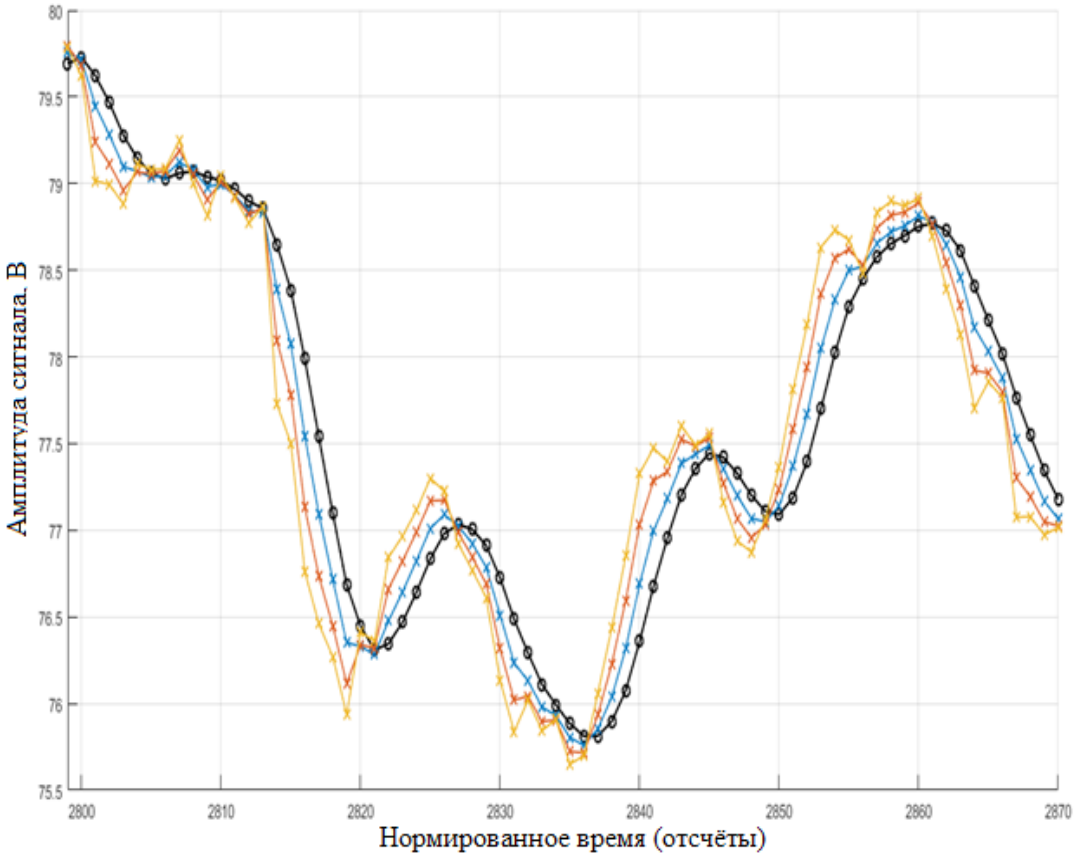


Рис. 7. Сформированный прогноз на один, два и три шага по времени вперед при наличии блока ЦОС и использовании дополнительных целевых отсчетов «стабилизаторов» (архитектура PAE)

В качестве примера на рис. 8 приведены зависимости ошибки прогноза на три шага вперед по времени для трех вариантов архитектур прогнозирующих модулей (рис. 1, 3, 4) (результаты тестирования, отображено в диапазоне от 2800 до 2870 нормированных временных отсчётов, то есть совпадает с рис. 7). Линия синего цвета – для архитектуры на рис. 1, линии красного и черного цвета для архитектур на рис. 3 и 4, соответственно.

Из рис. 8 видно, что линия (синяя линия) ошибки прогноза, соответствующая «полной» архитектуре прогнозирующего модуля (рис. 1) носит более сглаженный характер, что соответствует более высокому качеству результата прогнозирования.

В табл. 1 приведены рассчитанные значения СКО ошибки результата прогнозирова-

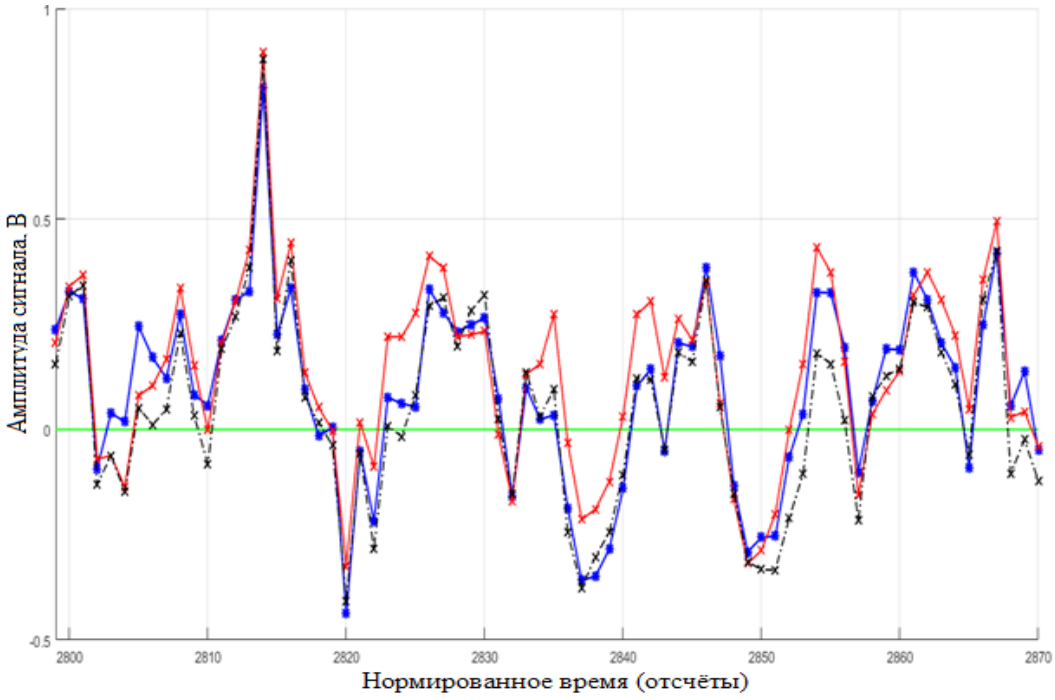


Рис. 8. Зависимости ошибки прогноза на три шага вперед по времени для трех вариантов архитектур прогнозирующих модулей (цвета линий: синий, красный и черный соответствуют рис. 1, 3, 4, соответственно)

ния на один, два и три шага по времени вперед для трех вариантов архитектур прогнозирующих модулей (рис. 1, 3, 4).

нений, собственно происходит обнаружение аномалий в наблюдаемых процессах АСУ ТП. При этом существенно возрастают требова-

Таблица 1

Среднеквадратические отклонение результатов прогноза для трех типов прогнозирующих модулей

Прогнозирующий модуль	СКО прогноза на 1 шаг вперед	СКО прогноза на 2 шага вперед	СКО прогноза на 3 шага вперед
Рисунок 3	0,0626	0,1463	0,2615
Рисунок 4	0,0401	0,1045	0,2026
Рисунок 1	0,0399	0,1044	0,1944

Из результатов, приведенным в табл. 1, следует, что архитектура прогнозирующего модуля (рис. 1), с использованием рассмотренного в работе прогнозирующего автоэнкодера (PAE) позволяет осуществлять прогнозирование временных рядов данных с более высокой точностью, что важно при использовании предложенного в работе прогнозирующего модуля в составе обнаружителя (детектора) аномалий процессов АСУ ТП.

Заключение

При формировании аномалий в потоках данных АСУ ТП будут происходить структурные изменения в сигнале ошибки формируемого прогноза наблюдаемого процесса АСУ ТП, по обнаружению этих структурных изме-

ния к точности настройки прогнозирующего модуля, формирующего эталонную модель процесса АСУ ТП, работающей в штатном режиме.

Показано, что предварительная ЦОС входного прогнозируемого сигнала в виде предложенного в работе блока из параллельного набора (гребенки) цифровых КИХ-ФНЧ, приводит к неравновесному учету корреляционных связей временных отсчетов входного сигнала и повышению точности результата прогнозирования.

Предложенный в работе прогнозирующий модуль, состоящий из последовательно соединенных блока ЦОС и прогнозирующего блока с использованием нейронной сети

(прогнозирующего автоэнкодера, predictive Autoencoder (PAE)), позволяет осуществлять прогнозирование временных рядов данных с более высокой точностью, что повышает качество обнаружения аномалий в процессах АСУ ТП.

Литература / References

1. Debar H., Dacier M., Wespi A. Towards a taxonomy of intrusion-detection systems // Computer Networks. 1999. vol. 31. Issue 8. pp. 805–822.

2. А. А. Браницкий, И. В. Котенко, Анализ и классификация методов обнаружения сетевых атак, Тр. СПИИРАН, 2016, выпуск 45, 207–244 DOI: https://doi.org/10.15622/sp.45.13 [A. A. Branitskiy, I. V. Kotenko, Analiz i klassifikatsiya metodov obnaruzheniya setevykh atak, Tr. SPIIRAN, 2016, vypusk 45, 207–244 DOI: https://doi.org/10.15622/sp.45.13].

3. Y. J. Xiao, W. Y. Xu, Z. H. Jia, Z. R. Ma & D. L. Qi (2017), «NIPAD: a Non-Invasive Power-based Anomaly Detection Scheme for Programmable Logic Controllers. Frontiers of Information Technology & Electronic Engineering», 18(4), 519-534.

4. W. Wang, Y. Xie, L. Ren, X. Zhu, R. Chang & Q. Yin (2018, May), «Detection of Data Injection Attack in Industrial Control System Using Long Short Term Memory Recurrent Neural Network», In 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA) (pp. 2710-2715), IEEE.

5. M. Kravchik & A. Shabtai (2018, October), «Detecting Cyber Attacks in Industrial Control Systems Using Convolutional Neural Networks», In Proceedings of the 2018 Workshop on Cyber-Physical Systems Security and PrivaCy (pp. 72-83), ACM.

6. P. Filonov, F. Kitashov & A. Lavrentyev (2017), «RNN-based Early Cyber-attack Detection for the Tennessee Eastman Process», arXiv preprint arXiv:1709.02232.

7. P. Filonov, A. Lavrentyev & A. Vorontsov (2016), «Multivariate Industrial Time Series with Cyber-attack Simulation: Fault Detection Using an Lstm-based Predictive Data Model», arXiv preprint arXiv:1612.06676.

8. A. N. Ragozin, V. F. Telezhkin, P. S. Podkorytov, «Forecasting Complex Multi-component Time Series within Systems Designed to Detect Anomalies in Dataflows of Industrial Automated Systems», SIN '19: Proceedings of the 12th International Conference on Security of Information and Networks, September 2019, Article No.: 2, pp. 1–5.

9. A. Ragozin, V.Telezhkin, P.Podkorytov, «Prediction of Aggregate Multicomponent Time Series in Industrial Automated Systems Using Neural Network», Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists 2019, 13-15 March, 2019, Hong Kong, pp. 17-19.

10. A. N. Ragozin, V. F. Telezhkin, P. S. Podkorytov. State Prediction in Control Systems via Compound Time Series: Neural Network Approach. EEE SoutheastCon 2019 Von Braun Center Huntsville, Alabama April 11th-14th, 2019. Pages 1–6.

11. S. Tomonobu and T. Hitoshi, «One-hour-ahead load forecasting using neural network», IEEE Transactions on Power Systems, no. 1, pp. 21–24, 2002.

12. R. Huang, T. Huang, and Gadh, «Solar generation prediction using the ARMA model in a laboratory-level micro-grid», in IEEE Third International Conference on Smart Grid Communications, Sydney, 2012, pp. 528–533.

13. A. M. Abdurakhmanov, M. V. Volodin, E. Yu. Zybin, and V. N. Ryabchenko, «Methods for predicting power consumption in distribution networks (review)», Russian Internet Journal of Electrical Engineering, pp. 3–23, 2016.

14. A. N. Ragozin, A. A. Razumov, «Neural network forecasting with preliminary digital filtering of complex radio signals», in XVI International scientific and technicalconference Physics and technicalapplications of wave processes, Miass, 2018, pp. 285–290.

15. M. Granroth-Wilding and S. Clark, «What Happens Next? Event Prediction Using a Compositional Neural Network Model», AAAI, pp. 2727–2733, 2016.

16. A. N. Sokolov, A. N. Ragozin, I. A. Pyatnitsky, S. K. Alabugin, «Applying of Digital Signal Processing Techniques to Improve the Performance of Machine Learning-based Cyber Attack Detection in Industrial Control System», SIN '19: Proceedings of the 12th International Conference on Security of Information and Networks, September 2019, Article No.: 23, pp. 1–4.

17. Muna, A. H., Moustafa, N., & Sitnikova, E. (2018). Identification of malicious activities in industrial internet of things based on deep learning models. Journal of Information Security and Applications, 41, 1-11

18. Hinton GE, Krizhevsky A, Wang SD. Transforming auto-encoders. In International Conference on Artificial Neural Networks 2011 Jun 14 (pp. 44-51). Springer, Berlin, Heidelberg.

19. Hinton, G. E., & Zemel, R. S. (1994). Autoencoders, minimum description length and Helmholtz free energy. In Advances in neural information processing systems 6 (pp. 3-10).

20. Sakurada, M., & Yairi, T. (2014, December). Anomaly detection using autoencoders with nonlinear dimensionality reduction. In Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis (p. 4). ACM

РАГОЗИН Андрей Николаевич, кандидат технических наук, доцент кафедры защиты информации, доцент кафедры инфокоммуникационных технологий высшей школы электроники и компьютерных наук ФГАОУ ВО «Южно-Уральский государственный университет (национальный исследовательский университет)». Россия, 454080, г. Челябинск, проспект Ленина, д. 76. E-mail: ragozinan@susu.ru

RAGOZIN Andrey, Candidate of Sciences in Technology, Department of Information Security, Department of Information Technology Federal State Autonomous Educational Institution of Higher Education «South Ural State University (national research university)» Russia, 454080, Chelyabinsk, prsp. Lenina, 76. E-mail: ragozinan@susu.ru

УДК 004.056.57 Вестник УрФО № 1(35) / 2020, с. 35–47

Гибилинда Р. В.

DOI: 10.14529/secur200104

КЛАСТЕРИЗАЦИОННЫЙ МЕТОД ИДЕНТИФИКАЦИИ ВОЗДЕЙСТВИЙ НА ФАЙЛЫ С ПРИМЕНЕНИЕМ АЛГОРИТМА К-СРЕДНИХ, ИСПОЛЬЗУЕМЫЙ ПРИ РАССЛЕДОВАНИИ ИНЦИДЕНТОВ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

В статье представлен кластеризационный метод идентификации воздействий на файлы, используемый при расследовании инцидентов информационной безопасности. Предлагаемый метод основан на применении алгоритма кластеризации k-средних с адаптированным алгоритмом автоматического определения оптимального количества кластеров, которые описывают воздействия на файлы. В статье рассмотрен процесс подготовки входных данных, полученных из записей журнала изменений тома $UsnJrnl, а также алгоритм выявления сложных комплексных воздействий на файлы, основанный на поиске взаимосвязей между кластерами. Предлагаемый кластеризационный метод имеет ярко выраженный автоматизированный характер, что позволяет специалисту, проводящему расследование инцидента информационной безопасности, ускорить процесс выявления и ликвидации последствий инцидента.

Ключевые слова: *расследование инцидентов информационной безопасности, воздействие на файл, кластеризация.*

Gibilinda R. V.

A CLUSTERING METHOD FOR IDENTIFYING FILE IMPACTS BASED ON THE K-MEANS ALGORITHM USED IN INFORMATION SECURITY INCIDENTS INVESTIGATION

The article presents a clustering method for identifying file impacts used in information security incidents investigation. The proposed method is based on application of k-means clusterization algorithm with adapted automatic optimal cluster number determination algorithm. Precisely defined clusters amount allows to group data to describe file impacts. The article discusses preparation process of input data obtained from $UsnJrnl volume changes log records, as well as the algorithm for identifying complex file impacts based on the search for relationships between clusters. The proposed clustering method has a pronounced automated character, which allows a specialist that carries out an information security incident investigation to speed up the process of identifying and eliminating the consequences of an incident.

Keywords: *information security incident response, file impact, clusterization.*

Развитие информационных технологий (ИТ) и укрепление их влияния на повседневную деятельность человеческого общества зачастую сопровождается повышением интереса злоумышленников к совершению противоправных действий в сфере ИТ. Рост объема информации, обрабатываемой в информационных системах (ИС), а также непрерывное развитие злоумышленниками методов и средств совершения киберпреступлений неизбежно приводят к необходимости постоянного совершенствования применяемых мер обеспечения информационной безопасности (ИБ).

Существующие методы и средства противодействия угрозам ИБ, направленные на обеспечение конфиденциальности, целостности и доступности информации, в силу ограниченности функционала, ошибок конфигурации и эксплуатации, не всегда могут полностью решить поставленные задачи по защите информации. Выявленные злоумышленниками недостатки применяемых методов и средств противодействия угрозам ИБ способствуют совершению киберпреступлений, результатом которых являются инциденты ИБ. Расследование инцидентов позволяет установить причины их возникновения, ликвидировать последствия, а также сформировать рекомендации по усовершенствованию мер обеспечения ИБ.

В процессе расследования инцидента ИБ с применением различных методов и средств [1] собирается и анализируется значительное количество данных, способствующих установлению причин инцидента, его последствий, а также действий злоумышленника: запуск и активность процессов, воздействия на файлы, установка сетевых соединений и передача данных и др. Указанные данные хранятся в нескольких источниках — «массивах данных»: журналы аудита [2], журналы событий операционной системы [3], записи о последних открытых файлах и запущенных программах, журналы средств защиты информации и др.

Одной из важных составляющих расследования инцидента ИБ является анализ массивов данных, в частности, идентификация воздействий на файлы. Это связано с тем, что

в ИС информация хранится в виде файлов. В рамках статьи под идентификацией воздействий на файлы будем понимать процесс, в результате которого определяется порядок изменения параметров, характеризующих файл.

Массивы данных, обработка и анализ которых лежат в основе процесса расследования инцидента ИБ, обладают рядом недостатков:

• отсутствие исчерпывающего и единоформатного набора параметров, характеризующих файл;

• возможность искажения данных массива даже в процессе функционирования операционной системы;

• отсутствие возможности автоматизированного анализа некоторых массивов данных.

Совокупность указанных недостатков порождает необходимость поиска новых методов и средств анализа данных, содержащихся в массивах, для достижения полноценного результата в расследовании инцидента ИБ. В работе [4] авторами рассмотрены проблемы формализации набора параметров, характеризующих файл, и верификации1 массива данных.

В тех ситуациях, когда данных в массиве немного, возможен ручной анализ. Тем не менее, анализ «вручную» обладает рядом недостатков:

• человеческий фактор – специалист, проводящий анализ массивов данных может упустить из виду информацию, связанную с произошедшим инцидентом;

• начало инцидента могло произойти значительно раньше, чем основная совокупность воздействий на файлы, обрабатываемые в ИС, в результате чего потенциальный «источник инцидента» не попадет в выборку по временному интервалу.

В работе [5] для идентификации воздействий на файлы использовался журнал изменений тома $UsnJrnl в качестве массива данных, который обладает рядом преимуществ:

• полнота — журнал содержит подробную информацию о действиях, совершенных по отношению к файлам;

• достоверность — системный файл $UsnJrnl защищен операционной системой от несанкционированных изменений со стороны пользователя;

• простота обработки данных, что играет немаловажную роль в последующей автоматизации их анализа.

В рамках настоящей статьи автором предлагается кластеризационный метод идентификации воздействий на файлы с использованием алгоритма k-средних при обработке записей журнала изменений тома $UsnJrnl.

Журнал $UsnJrnl представляет собой разреженный файл, расположенный в каталоге $Extend. Он состоит из двух потоков данных: $Max и $J. Первый содержит служебные данные о журнале и не представляет интереса в рамках расследования инцидента. Второй, напротив, является для специалиста, проводящего расследование инцидента ИБ, ценным источником информации, так как состоит из набора записей об изменениях, произошедших с файлами как в отношении содержимого, так и служебной информации [6].

Согласно [4], параметры, характеризующие файл j, представлены вектором:

$$V_j = \left\{ I_j, D_j, N_j, C_j, X_j \right\}, \qquad (1)$$

где:

• I_j – идентификатор файла – уникальное числовое значение, содержащееся в служебной информации о файле, используемое драйвером файловой системы для однозначного определения файла;

• D_j – идентификатор родительского каталога файла – уникальное числовое значение, используемое драйвером файловой системы для установления однозначного соответствия между файлом и каталогом, в котором файл расположен;

• N_j – имя файла – битовая строка, используемая драйвером файловой системы для представления файла пользователю;

• C_j – содержимое файла – битовая строка, являющаяся информацией, хранимой в файле;

• X_j – иная служебная информация о файле – набор числовых значений, являющихся служебной информацией о файле, зависящий от типа файловой системы.

Записи журнала $UsnJrnl содержат в явном виде только часть параметров (компонентов вектора V_j), тем не менее, поле записи R_j может быть косвенно использовано для определения изменений в компонентах C_j и X_j в рамках решения задачи верификации воз-

[1] В рамках статьи под верификацией массива данных будем понимать процесс выявления комбинаций параметров, характеризующих файл, возникновение которых невозможно в процессе штатного заполнения массива данными.

действия на файл. Формат записи представлен в таблице 1. Курсивом выделены поля, информация в которых позволяет идентифицировать файл и действия, осуществленные по отношению к нему. Полужирным шрифтом выделены поля, которые являются основой для предлагаемого автором кластеризационного метода.

Таблица 1

Формат записи журнала изменений тома $UsnJrnl

Смещение, байт	Размер, байт	Описание
0x00	4	Размер записи
0x04	2	Версия записи
0x06	2	Версия программного обеспечения, которым запись создана
0x08	8	Идентификатор файловой записи (Ij)
0x10	8	Идентификатор родительского каталога (Dj)
0x18	8	**Номер записи (Uj)**
0x20	8	**Временная отметка создания записи (T)**
0x28	4	Идентификатор действия (Rj)
0x2C	4	Тип источника записи
0x30	4	Идентификатор безопасности
0x34	4	Атрибуты файла
0x38	2	Длина имени файла *
0x3A	2	Начало имени файла в записи
0x3C	*	Имя файла (Nj)

В рамках настоящего исследования был проанализирован порядок формирования записей и выявлена зависимость, описываемая выражением:

$$\Psi = f(\tau, K_j), \qquad (2)$$

где Ψ – множество значений номеров записей журнала изменений тома $UsnJrnl, имеющих отношение к файлу *j*, причем $\forall j\, U_j \in \Psi$; K_j – количество записей журнала, имеющих отношение к файлу *j*; τ – множество временных интервалов различной величины, в рамках которых осуществляются действия по отношению к файлу *j*, описываемые записями журнала изменений тома $UsnJrnl, причем $T \in \tau$.

Пример зависимости (2), аппроксимированный кусочно-заданной функцией, представлен на рис. 1. По оси абсцисс отложены отнормированные к минимальному значения *T*, а по оси ординат – отнормированные к минимальному значения U_j нескольких записей журнала $UsnJrnl. Окружностями обозначены значения (T, U_j) каждой записи для файла *j*.

Выявленная зависимость обладает следующими свойствами:

• нелинейная – за фиксированный интервал времени может быть совершено произвольное количество воздействий на файлы, что повлечет за собой появление соответствующего количества записей журнала $UsnJrnl;

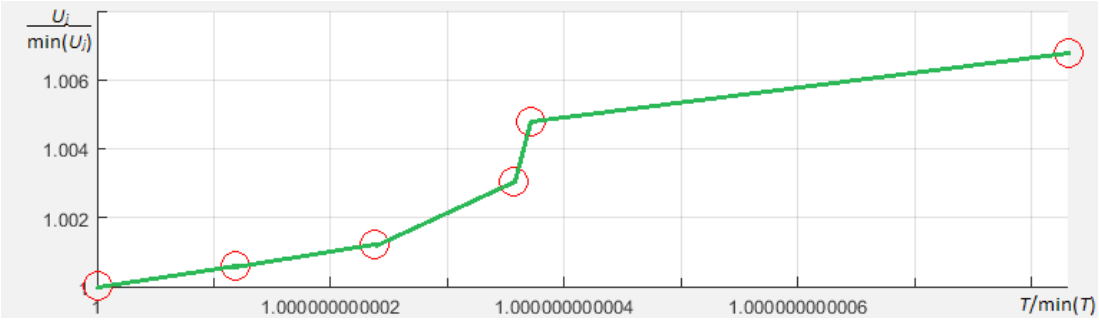


Рис. 1. Зависимость отнормированных к минимальному значений Uj от количества записей журнала $UsnJrnl в нескольких отнормированных к минимальному временных интервалах

• возрастающая – номер записи постоянно увеличивается. Несмотря на наличие ограничения в размерности поля U_j (8 байт), ис-

черпать его ресурс при повседневной активности пользователя затруднительно.

Нелинейность позволяет определить период наибольшей активности пользователя/процесса по отношению к файлу. Крутизна графика функции в выбранном фиксированном временном интервале характеризует интенсивность воздействий на файлы. Рост значения номера записей U_j говорит об интенсивности работы с файлами в целом. При совершении простых воздействий крутизна графика функции незначительна. К ним можно отнести работу с документами в формате txt, простые файловые операции, такие как создание/удаление/переименование файла. Сложные комплексные воздействия на файлы существенно увеличивают крутизну графика. Примерами могут служить редактирование «офисного» документа в текстовом процессоре Microsoft Word, разархивирование файлов и т. д. Изменяемое значение крутизны графика и нелинейность позволяют рассмотреть использование математических алгоритмов для проведения анализа выявленной зависимости.

Предлагаемый метод базируется на реализации процесса автоматизированного объединения нескольких записей в совокупность с целью получения воздействия на файл. Для этого возможно использование различных типов алгоритмов машинного обучения: кластеризации, классификации, нейронных сетей и др. В рамках исследования были рассмотрены алгоритмы кластеризации *k*-средних, *k*-медоидов, DBSCAN.

DBSCAN [7], несмотря на возможность работы с кластерами произвольной формы, требует точного задания двух параметров: ε и minPts. В случае неоптимального задания этих параметров данные либо не будут кластеризованы, либо сольются в один кластер. Алгоритм показал отрицательные результаты при группировке данных с различной плотностью распределения в пространстве. Модификации алгоритма, например, Generalized DBSCAN или Hierarchical DBSCAN существенно уступают в скорости работы по сравнению с *k*-средних и *k*-медоидов в рамках решаемой задачи.

Алгоритм *k*-медоидов похож на *k*-средних как по принципу работы, так и по скорости выполнения, отличие заключается лишь в порядке выбора центра кластера – центром выбирается одна из точек кластера, а не его «центр масс». В связи с особенностью выбора центра кластера, алгоритм *k*-медоидов показал отрицательные результаты в тех случаях, когда точки в пространстве распределены в формах половины и четверти окружности, что не является редкостью в связи с особенностями зависимости (2).

Для автоматизации процесса идентификации воздействий на файлы в рамках настоящего исследования был выбран алгоритм кластеризации *k*-средних по соотношению скорость работы/сложность подготовки входных данных/эффективность полученных результатов.

Объем записей в журнале $UsnJrnl может превышать 600 тысяч, что, в свою очередь, затрудняет использование алгоритма *k*-средних для группировки записей в рамках предлагаемого метода идентификации воздействий на файлы при расследовании инцидентов информационной безопасности – продолжительность анализа записей может достигать нескольких часов в зависимости от конфигурации компьютера. Для повышения эффективности применения алгоритма *k*-средних в отношении решаемой задачи необходимо подготовить входные данные, выбрать оптимальное количество кластеров, которые представляют собой воздействия на файлы, и проанализировать связи между кластерами для выявления сложных комплексных воздействий.

Для повышения скорости анализа автором предлагается использовать в качестве входных данных алгоритма *k*-средних «порции» записей из $UsnJrnl, предварительно разделенных согласно идентификаторам файлов I_j. Помимо выборки данных по I_j необходимо также учесть особенность работы драйвера файловой системы (ФС) NTFS, а именно активное использование им существующих освобожденных файловых записей вместо выделения новых. Указанная особенность может привести к получению неверного результата.

Для предотвращения ошибок анализа записей журнала $UsnJrnl из-за особенностей работы драйвера ФС NTFS при разделении записей по идентификаторам файлов I_j предлагается алгоритм разделения записей журнала $UsnJrnl на блоки (алгоритм № 1):

1. Выбрать записи в соответствии с идентификатором I_j.

2. Создать буфер «разделенных» записей – буфер 1 и назначить его текущим.

3. Провести в цикле последовательную

проверку выбранных записей на наличие значения R_j = 0x00000200, свидетельствующего об удалении файла.

4. Если значение 0x00000200 в поле R_j записи не найдено, то поместить запись в текущий буфер.

5. В противном случае создать следующий буфер (буфер 2), назначить его текущим. Если значение 0x00000200 в поле R_j записей появляется m раз, то создать m+1 буферов, чтобы разделить записи на несколько частей в соответствии с особенностями работы драйвера ФС NTFS.

6. Поместить запись в текущий буфер.

7. Разделенные на части записи сохранить в виде блоков для последующего анализа.

Схема алгоритма представлена на рис. 2.

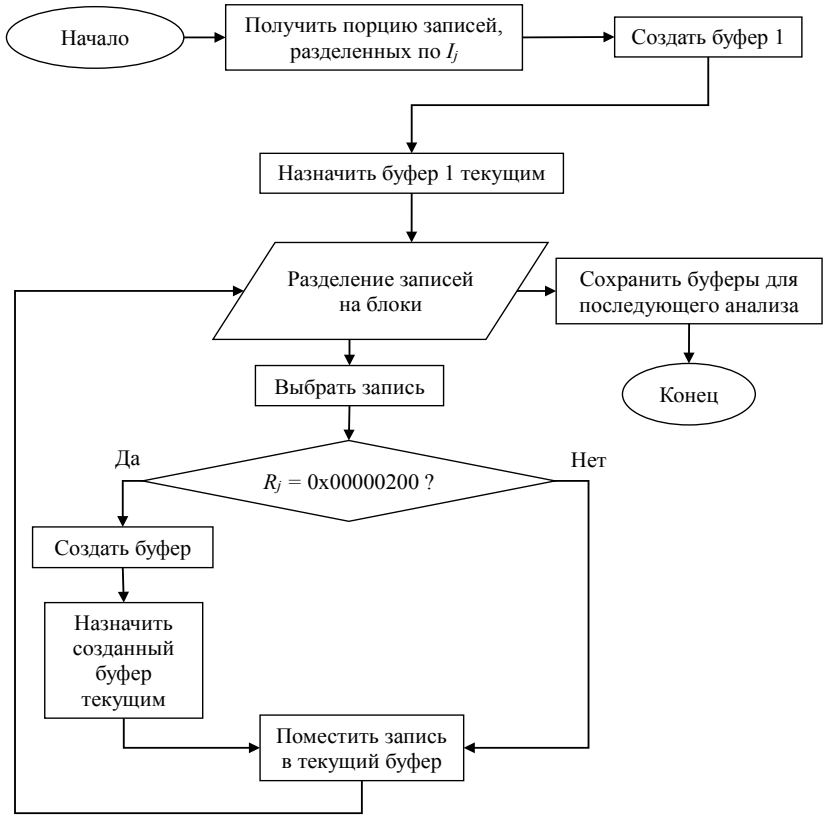


Рис. 2. Алгоритм №1 для учета особенностей драйвера ФС NTFS

После разделения записей в связи с особенностями работы драйвера ФС NTFS значения полей U_j и T каждой разделенной части следует отнормировать к соответствующему минимуму. Нормировка необходима для того, чтобы проводить анализ зависимости (2) в числовых значениях одного порядка.

Необходимым условием работы алгоритма *k*-средних является указание желаемого количества кластеров *k* для группировки входного множества значений. От выбора значения *k* зависит корректность получаемых результатов. Пример задания верного количества кластеров указан на рис. 3, где видно, что точки в пространстве были сгруппированы и обозначены цветом оптимальным образом. При установке значения *k*, отличного от 3, для группировки тех же данных, полученный результат окажется некорректным.

Существуют несколько методов, позволяющих определить оптимальное значение *k*: метод «локтя», удовлетворение информационным критериям Акаике или Шварца, метод «силуэтов», построение дендрограммы кластеризации, метод градиентного спуска и др. [8]. Описание достоинств и недостатков указанных методов выходит за рамки данной статьи. Для решения задачи выбора оптимального значения *k* выбран принцип минимальной длины описания (MDL) [9–11].

Принцип MDL основывается на следующей идее: «Любая закономерность в задан-

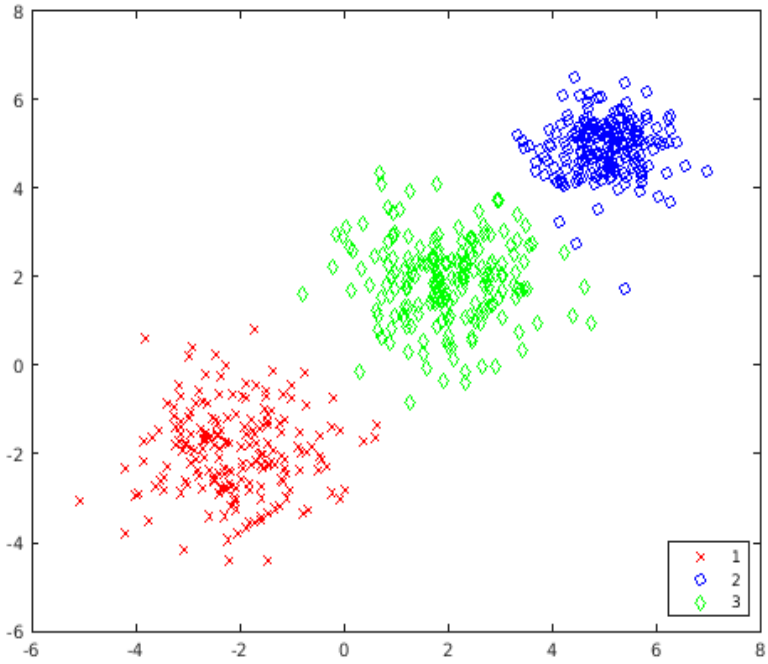


Рис. 3. Разделение точек на плоскости при значении $k = 3$

ном наборе данных может быть использована для сжатия данных, то есть описания данных с использованием меньшего набора символов, чем нужно для описания данных буквально» [11]. В роли количественной оценки величины, которая необходима для описания набора данных, выступает длина описания L(x). В соответствии с принципом MDL, нахождение минимального значения L(x) позволяет определить оптимальный способ описания данных. В статье автор предлагает использовать принцип MDL и нахождение минимального значения L(x), то есть, в отношении решаемой задачи, оптимального количества кластеров k, которые описывают воздействия на файлы.

В работе [12] для расчета длины описания исследуемых случайно распределённых значений предлагается использовать следующую формулу:

$$L(x) = -\sum_{x \in X} \log_2 p(x) + \frac{1}{2} P \log_2 |X|, \quad (3)$$

где X – вектор исследуемых значений, *p(x)* – плотность вероятности исследуемого значения *x*, *P* – количество переменных параметров в формуле плотности вероятности.

В формуле (3) длина описания складывается из «стоимостей» описания. Первое слагаемое является логарифмической функцией правдоподобия исследуемых значений и описывает «стоимость» описания этих значений. Второе слагаемое представляет собой «стоимость» описания мощности множества значений, заданных в соответствии с предполагаемым/выбранным законом плотности вероятности.

Формула (3) была адаптирована в работе [13] для поиска оптимального значения кластеров в алгоритме k-медоидов. Адаптированная формула описывается выражением:

$$L(x) = -\sum_{x \in X} \log_2 p(\|x - c_x\|) + (\frac{1}{2} P + k) \log_2 |X|, \quad (4)$$

где X – вектор элементов исследуемых значений, $p(\|x - c_x\|)$ – плотность вероятности расстояния между исследуемым значением x (точки с заданными координатами) и ближайшим к нему центром кластера c_x, *P* – количество переменных параметров в формуле плотности вероятности, *k* – количество кластеров.

Формула (4) может быть использована для расчета минимальной длины описания в целях определения оптимального значения k. По мнению автора статьи, формула (4) обладает избыточностью, которая усложняет решение задачи определения оптимального значения k для алгоритма k-средних.

В ходе настоящего исследования проведен анализ влияния слагаемых формулы (4) на определение оптимального значения k в рамках предлагаемого метода и получены следующие результаты:

1. Данные, которые попадают на вход алгоритма *k*-средних, предварительно разделяются на «порции», что существенно уменьша-

ет абсолютную величину слагаемого $\frac{1}{2}P\log_2|X|$ и его вклад в значение длины описания.

2. Абсолютное значение длины описания *L(x)* не важно для определения оптимального количества кластеров, т.к. из полученных значений *L(x)* выбирается минимальное.

3. В ходе экспериментов расчет *L(x)* проводился для нескольких законов плотности вероятности (Probability Density Function – PDF), в отличие от работы [13]. Необходимость выбора из нескольких PDF обусловлена тем, что при расчете *L(x)* в рамках предлагаемого метода анализа записей журнала $UsnJrnl предполагаемое распределение величин расстояний между значениями *x* (точек с координатами *(T, U_j)*) и центрами ближайших кластеров c_x (рассчитанных алгоритмом *k*-средних точек с координатами (*T', U_j'*)) задается специалистом. Влияние на распределение величин расстояний оказывают типы воздействий на файлы, а также интенсивность воздействий. Результаты экспериментов по выбору наилучших значений параметров PDF представлены в таблице 2.

ных экспериментов, автором предлагается адаптированная формула для расчета длины описания, используемая при определении количества кластеров *k*:

$$L(x) = -\sum_{m=1}^{M}\log_2 p(\|x_m - c_m\|) + k\log_2 M \ , \quad (5)$$

где M – количество записей журнала $UsnJrnl с отнормированными значениями *T* и U_j в «порции», $p(\|x_m - c_m\|)$ – плотность вероятности расстояния между исследуемым значением xm (точки с координатами $\langle T, U_j \rangle_m$) и ближайшим к нему центром кластера cm (точки с рассчитанными координатами $\langle T', U_j' \rangle_m$).

Для определения оптимального количества кластеров с учетом формулы (5) и проведения кластеризации «порции» записей журнала $UsnJrnl необходимо воспользоваться алгоритмом определения значения *k* (алгоритм № 2):

1. Выбрать «порцию» записей журнала $UsnJrnl, предварительно обработанных с использованием алгоритма № 1.

2. Выбрать PDF с оптимальными параметрами согласно таблице 2.

Таблица 2

Значения параметров PDF, используемых при определении оптимального количества кластеров

PDF	Параметр	Диапазон значений параметра / Наилучшее значение параметра
Нормальное распределение	μ	0-2 / 0
	σ	0.5-1.5 / 1
Гамма-распределение	α	0.5;1;2 / 1
	β	0.8-1.1 /0.9
Распределение Стьюдента	ν	1-5 / 1
Распределение χ^2	ν	1-5 / 2
Распределение Фишера	$\nu 1$	1-2 / 1
	$\nu 2$	1-5 / 2

В ходе экспериментов было установлено, что для идентификации большинства воздействий на файлы достаточно использовать PDF нормального распределения с оптимальными значениями параметров, указанными в таблице 2, при расчёте значения *L(x)* для определения оптимального количества кластеров *k*.

С учетом анализа результатов проведен-

3. Определить количество записей *M* в «порции».

4. Запустить цикл с изменением значения *k* от 1 до *M*. На каждой итерации цикла:

a. С использованием алгоритма *k*-средних для текущего значения *k* вычислить центры кластеров cm текущей «порции» записей;

b. С использованием формулы (5) рассчитать значение *L(x)*;

c. Если значение *L(x)* текущей итерации меньше или равно значению, полученному на предыдущей итерации, то продолжить выполнение цикла. В противном случае прервать выполнение – минимальное значение *L(x)* найдено.

d. Определить оптимальное значение *k*, соответствующее минимальному *L(x)*.

тате проведенной кластеризации «порции» записей журнала $UsnJrnl выделены два воздействия на файл *j*, состоящие из 12 и 4 записей журнала изменений тома $UsnJrnl соответственно. Центры кластеров, описывающих выделенные воздействия, обозначены плюсами.

Полученные в ходе применения алгорит-

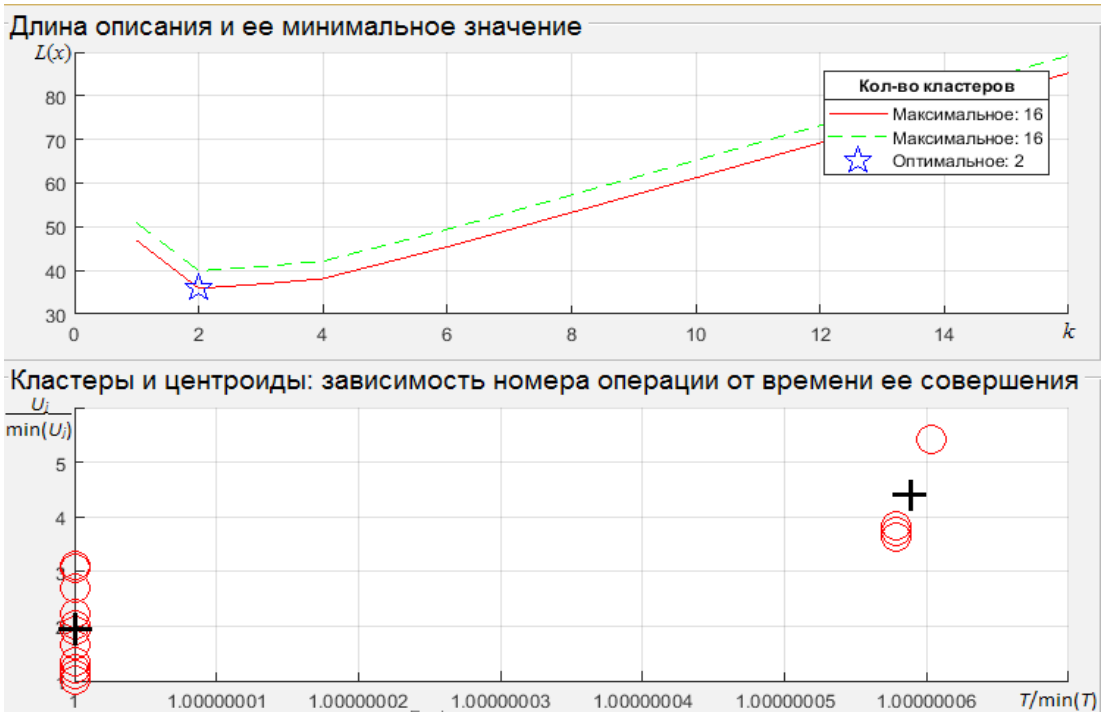


Рис. 4. Пример определения оптимального количества кластеров для идентификации воздействия на файл

5. Сохранить результаты кластеризации проведенной «порции» записей журнала $UsnJrnl при оптимальном значении *k*.

Пример определения оптимального количества кластеров и кластеризации записей журнала $UsnJrnl показан на рис. 4. На верхней оси координат: пунктирной линией показан график значений *L(x)*, рассчитанных для нескольких значений *k* по формуле (4); сплошная линия показывает график значений *L(x)*, рассчитанных по формуле (5); звездой отмечено оптимальное значение *k*. На нижней оси координат: окружностями обозначены значения xm – точки с координатами $\langle T, U_j \rangle_m$, полученные из записей журнала $UsnJrnl, плюсами – центры кластеров c_m. Примечание: для повышения наглядности результатов, выполнение цикла алгоритма № 2 не прерывалось при нахождении минимального значения *L(x)*.

Как видно из графиков верхней части рисунка, слагаемое $\frac{1}{2}P\log_2|X|$ не вносит существенного влияния в значение *L(x)*. В результате

мов № 1 и № 2 результаты свидетельствуют о том, что объединение записей журнала $UsnJrnl в кластеры, описывающие события над файлами, сокращает, в некоторых случаях существенно, объем анализируемых данных в ходе расследования инцидента ИБ.

При предварительной подготовке данных алгоритмом № 1 было сделано допущение о возможности разделения записей журнала $UsnJrnl по идентификаторам файловых записей. Такое допущение позволило существенно уменьшить время обработки данных с применением алгоритма *k*-средних, но в то же время снизило точность идентификации сложных комплексных воздействий, состоящих из нескольких записей журнала $UsnJrnl, в том числе имеющих отношение к различным файлам. Для того, чтобы корректно идентифицировать сложные комплексные воздействия, необходимо проанализировать связи между полученными кластерами.

Для проведения анализа взаимосвязей между кластерами в процессе идентифика-

ции сложных комплексных воздействий необходимо последовательно обработать все кластеры в целях поиска файлов с одинаковыми именами с использованием алгоритма № 3:

1. Определить полученное количество кластеров по результатам работы алгоритмов № 1 и № 2.

2. В цикле осуществить последовательный перебор кластеров. На каждой итерации цикла:

a. Выбрать кластер, извлечь из его элементов (записей журнала $UsnJrnl) все имена файлов;

b. Произвести поиск совпадений имен файлов в остальных кластерах;

c. Отсортировать кластеры по значению поля T записи журнала $UsnJrnl, расположенной в начале каждого кластера;

d. Последовательно объединить кластеры, если временной интервал между конечным элементом текущего кластера и начальным элементом следующего не превышает значения *t1*;

e. Предложить объединение кластеров, если временной интервал между конечным элементом текущего кластера и начальным элементом следующего не превышает значения *t2*;

f. Пометить все кластеры, которые участвовали в объединении на данной итерации цикла, как обработанные.

3. Вывести предложения об объединении кластеров, полученные на шаге 2.e алгоритма, специалисту, проводящему расследование инцидента ИБ.

Значения параметров *t1* и *t2* задаются в рамках предложенных значений специалистом, проводящим расследование инцидента ИБ, исходя из ожидаемой точности объединения кластеров записей журнала $UsnJrnl: чем больше значение *t1* и *t2*, тем больше кластеров будет объединено. Диапазоны рекомендуемых значений параметров:

• *t1*: 200 – 800 мс, оптимальное значение 400 мс;

• *t2*: 2 – 8 с, значение по-умолчанию 3 с.

Интерфейс программного обеспечения с предложением об объединении кластеров представлен на рис. 5.

На представленном рисунке столбцы таблиц соответствуют следующим полям записей журнала $UsnJrnl:

• столбец «ИФЗ» содержит идентификатор файла I_j, равный 43 и 44 в двух кластерах;

• столбец «ИРК ФЗ» содержит идентификатор родительского каталога D_j, равный 5 в обоих кластерах;

• столбец «Файловая операция» содержит идентификаторы действий R_j, совершенных по отношению к файлам;

• столбец «Имя файла» содержит имена файлов N_j;

• столбец «USN» содержит номера записей U_j.

Из рисунка видно, что в процессе объединения кластеров записи журнала $UsnJrnl о двух файлах с идентификаторами I_j, равными 43 и 44, попадут в объединенный кластер и будут принадлежать одному сложному комплексному воздействию на файл.

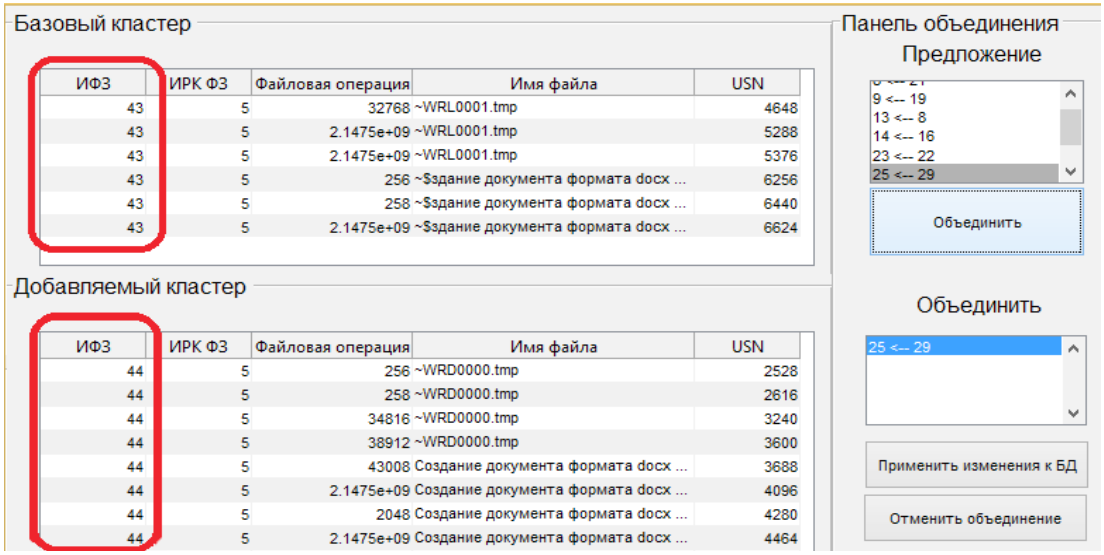


Рис. 5. Интерфейс программного обеспечения с предложением по объединению кластеров

Полученные с применением алгоритмов №№ 1 – 3 воздействия на файлы позволяют сократить объем обрабатываемой в рамках расследования инцидента ИБ информации. В ходе анализа может возникнуть вопрос достоверности полученных воздействий на файлы с позиций корректности результатов работы алгоритмов и отсутствия искажений в массиве данных. Не смотря на то, что используемый массив данных – журнал изменений тома $UsnJrnl защищен от изменений средствами операционной и файловой систем, существует возможность внесения искажений в массив с применением загрузки с внешнего носителя.

Верифицировать полученные воздействия на файлы возможно с применением событийной модели [4], на вход которой подаются:

• идентификаторы файла I_{j1} и I_{j2};

• идентификаторы родительского каталога D_{j1} и D_{j2};

• имена файла N_{j1} и N_{j2};

• признак изменения содержимого Z.

Для функционирования модели должна быть выставлена начальная маркировка μ, соответствующая равенству заданных входных параметров и установке признака изменения содержимого. Параметр Z можно определить по таблице, указанной в работе [5], исходя из значения поля R_j записи, принадлежащей верифицируемому кластеру. Параметр τ вычисляется как разница между значениями полей T смежных записей журнала $UsnJrnl, принадлежащих верифицируемому кластеру.

В результате проведения верификации специалист, проводящий расследование инцидента ИБ, проверяет последовательность записей в кластере на предмет наличия в них искажений. Стоит отметить, что записи в журнале $UsnJrnl формируются в соответствии с собственным алгоритмом драйвера файловой системы NTFS и не зависят от особенностей работы событийной модели. В связи с этим, может возникнуть несоответствие набора входных параметров искомому воздействию на файл в случае сложного комплексного воздействия. Специалист может убедиться в отсутствии искажений путем проверки последовательности значений полей U_j – номера записей при переходе между файлами в рамках одного воздействия либо смежные, либо отличаются на 1-2 записи.

Предлагаемый кластеризационный метод идентификации базируется на последовательном применении рассмотренных алгоритмов №№ 1 – 3. Для идентификации воздействий на файл предложенным методом необходимо последовательно выполнить несколько шагов:

1. Считать массив данных – журнал изменений тома $UsnJrnl, сохранив содержащиеся в нем записи в базу данных.

2. Создать множество выборок данных, разделив их по идентификаторам файлов I_j;

3. К каждой выборке данных применить алгоритм № 1, чтобы подготовить данные для проведения кластеризации алгоритмом *k*-средних;

4. Кластеризовать каждую подготовленную выборку алгоритмом *k*-средних, определив оптимальное количество кластеров k в соответствии с алгоритмом № 2;

5. Сохранить кластеры, описывающие каждую выборку, в базу данных;

6. Применить алгоритм № 3 в целях поиска взаимосвязей между кластерами в базе данных для повышения точности идентификации сложных комплексных воздействий. Применить/отклонить предложения по объединению кластеров, как результат работы алгоритма № 3.

Верифицировать кластеры с применением событийной модели.

Сформировать список идентифицированных воздействий на файлы, описанных кластерами записей журнала $UsnJrnl.

В сформированном списке найти те воздействия на файлы, которые имеют отношение к расследуемому инциденту ИБ.

Примеры полученных с помощью предлагаемого метода воздействий представлены на рис. 6 и 7.

На рис. 6 представлено воздействие на файл, в рамках которого создается файл taskhsvc.exe. Отличительным свойством указанного воздействия является изменение содержимого исполняемого файла (значения 2 и 3 в столбце «Файловая операция»), что может быть интерпретировано как активность вредоносного программного обеспечения, так и процедура обновления программного обеспечения. Для установления факта, свидетельствующего об активности вредоносного программного обеспечения, следует проанализировать остальные идентифицированные воздействия. Например, воздействие, представленное на рис. 7 говорит о том, что операционная система подверглась заражению

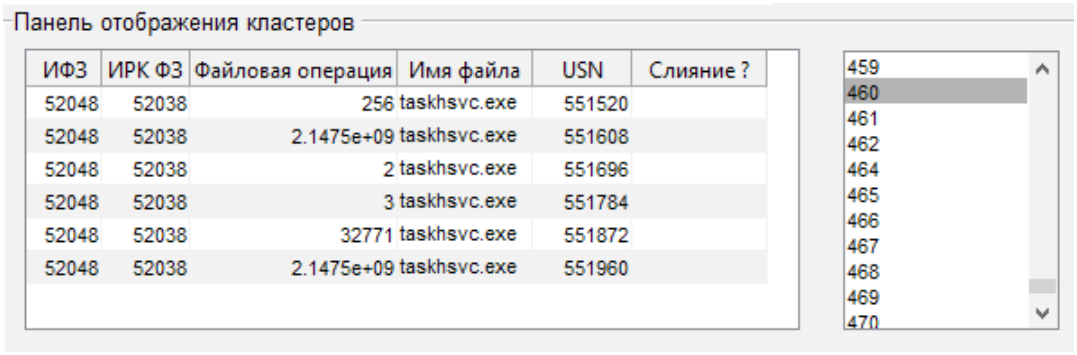


Рис. 6. Пример воздействия: создание исполняемого файла и изменение его содержимого

Рис. 7. Пример воздействия: создание файла с изображением в формате bmp

вредоносным программным обеспечением WannaCry, так как создается файл с характерным именем.

При рассмотрении нескольких воздействий специалист, проводящий расследование инцидента ИБ, может сделать вывод о том, что файл, созданный в рамках идентифицированного на рис. 6 воздействия, является копией WannaCry, замаскированной под штатный сервис операционной системы. Последующий анализ остальных воздействий на файлы поможет выявить «зараженные» файлы, и определить стратегию ликвидации последствий инцидента ИБ.

Предлагаемый кластеризационный метод позволяет идентифицировать воздействия на пользовательские и системные файлы, определять попытки внесения изменений в исполняемые файлы вредоносным программным обеспечением и т.д. Идентификация воздействий на файлы позволяет ускорить процесс расследования инцидента ИБ, а также упростить ликвидацию его последствий.

Метод имеет ярко выраженный автоматизированный характер, что способствует созданию на его основе программного обеспечения, позволяющего специалисту, проводящему расследование инцидента ИБ, сконцентрировать свое внимание на установлении причин инцидента, установлении взаимосвязей между его событиями, нежели на рутинном поиске информации о событиях инцидента. Особенностью метода является возможность проведения верификации анализируемых данных с целью установления попыток их искажения.

Литература / References

Стандарт Банка России СТО БР ИББС-1.3-2016 «Обеспечение информационной безопасности организаций банковской системы Российской Федерации. Сбор и анализ технических данных при реагировании на инциденты информационной безопасности при осуществлении переводов денежных средств» [Электронный ресурс]. URL: http://garant.ru/products/ipo/prime/doc/71457690 (дата обращения: 10.03.2020). [Standart Banka Rossii STO BR IBBS-1.3-2016 «Obespechenie informatsionnoy bezopasnosti organizatsiy bankovskoy sistemy Rossiyskoy Federatsii. Sbor i analiz tekhnicheskikh dannykh pri reagirovanii na intsidenty informatsionnoy bezopasnosti pri osushchestvlenii perevodov denezhnykh sredstv» [Elektronnyy resurs]. URL: http://garant.ru/products/ipo/prime/doc/71457690 (data obrashcheniya: 10.03.2020)].

1. Studiawan H., Payne C., Sohel F. Graph Clustering and Anomaly Detection of Access Control Log for Forensic Purposes // Digital Investigation (2017). 2017.

2. Dwyer J., Marius Truta T. Finding Anomalies in Windows Event Logs Using Standard Deviation // 9th IEEE International on Collaborative Computing: Networking, Applications and Worksharing. 2013. P. 563 -570.

3. Гайдамакин Н. А., Гибилинда Р. В., Синадский Н. И. Событийная модель процесса идентификации воздействий на файлы при расследовании инцидентов информационной безопасности, основанная на математическом аппарате сетей Петри // Вестник СибГУТИ. 2020. № 1. (в печати).

Gaidamakin N., Gibilinda R., Sinadskiy N. File operations information collecting software package used in the information security incidents investigation // IEEE Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT-2020). Yekaterinburg, Russia, May 14 – 15, 2020. (in press). [Gaydamakin N. A., Gibilinda R. V., Sinadskiy N. I. Sobytiynaya model' protsessa identifikatsii vozdeystviy na fayly pri rassledovanii intsidentov informatsionnoy bezopasnosti, osnovannaya na matematicheskom apparate setey Petri // Vestnik SibGUTI. 2020. № 1. (v pechati)].

4. USN_RECORD_V2 – Win32 apps. [Электронный ресурс]. URL: https://docs.microsoft.com/en-us/windows/win32/api/winioctl/ns-winioctl-usn_record_v2 (дата обращения: 10.03.2020).

5. Ester M., Kriegel H., Sander J., Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise //Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. AAAI Press. 1996. P.226-231.

6. Determining the number of clusters in a data set. [Электронный ресурс]. URL: https://en.wikipedia.org/wiki/Determining_the_number_of_clusters_in_a_data_set (дата обращения: 10.03.2020).

7. Rissanen J. Modeling by shortest data description // Automatica, Vol. 14 (1978), pp. 465-471.

8. Rissanen J. A Universal Prior for Integers and Estimation by Minimum Description Length // The Annals of Statistics, vol. 11 (1983), no. 2, pp. 416-431.

9. Grunwald P. D., Myung J. I., Pitt M. A. Advances in Minimum Description Length: Theory and Applications. Cambridge, Massachusetts; London, England: MIT Press, 2005. 372p.

10. Roberts S. Novelty detection using extreme value statistics // IEE Proceedings – Vision, Image and Signal Processing, vol. 146 (1999), no. 3, pp.124-129.

11. Using minimum description length to optimize the 'k' in k-medoids [Электронный ресурс]. URL: http://erikerlandson.github.io/blog/2016/08/03/x-medoids-using-minimum-description-length-to-identify-the-k-in-k-medoids (дата обращения: 10.03.2020).

ГИБИЛИНДА Роман Владимирович, ассистент учебно-научного центра «Информационная безопасность», Институт радиоэлектроники и информационных технологий - РТФ Уральский федеральный университет им. первого Президента России Б.Н. Ельцина. 620002, г. Екатеринбург, ул. Мира, 19. E-mail: r.v.gibilinda@urfu.ru

GIBILINDA Roman, assistant of Educational and Scientific Center "Information Security", Institute of Radio electronics and Information Technologies, Ural Federal University named after first President of Russia B.N. Yeltsin. 620002, Yekaterinburg, Mira str., 19. E-mail: r.v.gibilinda@urfu.ru

УДК 004.056 Вестник УрФО № 1(35) / 2020, с. 48–57

Домуховский Н. А., Синадский А. Н.

DOI: 10.14529/secur200105

ИТЕРАТИВНЫЙ СТАТИСТИКО-ЭНТРОПИЙНЫЙ МЕТОД И АЛГОРИТМ АНАЛИЗА СЕТЕВОГО ТРАФИКА ПРИ ОТСУТСТВИИ АПРИОРНЫХ СВЕДЕНИЙ О ЕГО СТРУКТУРЕ

Статья посвящена анализу трафика при отсутствии априорных сведений о его структуре с целью выявления уязвимостей и проведения аудита. В результате объединения существующих энтропийного и статистического алгоритмов разработан статистико-энтропийный метод выделения сетевых узлов и значимых полей из трафика неизвестных протоколов. Энтропийный алгоритм, анализируя массив трафика, на основе энтропии отдельных байт и взаимной информации пар байт принимает решение о границах значимых полей. Статистический алгоритм для определения сетевых адресов использует оценку количества вхождений похожих на части сетевого пакета подстрок в ранее полученный массив сетевого трафика. На основе энтропийного алгоритма разработан итеративный алгоритм, решающий задачу анализа трафика, имеющего в своём составе более одного протокола. Математические модели каждого из алгоритмов реализованы программно, результатом работы программной реализации описанного статистико-энтропийного метода из сетевого трафика без априорных сведений об используемых в нём протоколах выделяются сетевые адреса и предлагается разделение на семантические поля.

Ключевые слова: *анализ сетевого трафика, реверс-инжиниринг, статистика.*

Domukhovsky N.A., Sinadsky A.N.

ITERATIVE STATISTICAL-ENTROPY METHOD FOR ZERO KNOWLEDGE NETWORK TRAFFIC ANALYSIS ALGORITHM IMPLEMENTATION

The article is devoted to traffic analysis with zero knowledge about its structure. As a result of combining existing entropy and statistical algorithms, a statistical-entropy method has been developed capable of distinguishing network nodes and significant fields from traffic with un-

known protocol. The decision about significant fields boundaries in the analyzed traffic sample made by the algorithm is based on the entropy of individual bytes and byte pairs mutual information. The statistical algorithm determines network addresses using estimate number of occurrences parts of a network packet similar (as a strings) to parts of a previously received array of network traffic. Based on the entropy algorithm, an iterative algorithm has been developed that solves the problem of traffic analysis, which includes more than one protocol. The mathematical models each of the algorithms are implemented as a module of the program that implements the statistical-entropy method. As a result of the software implementation of the described statistical-entropy method, network addresses are allocated from the network traffic with zero knowledge about the protocols used in it, and separation into semantic fields is proposed.

Keywords: *network traffic analysis, reverse engineering, statistics.*

В Требованиях по обеспечению безопасности значимых объектов критической информационной инфраструктуры Российской Федерации, утвержденных приказом ФСТЭК России от 25 декабря 2017 г. №239, в качестве меры АУД.5 указан «Контроль и анализ сетевого трафика». При проведении мониторинга в условиях проприетарных протоколов, распространённых как в сетях АСУ ТП, так и IoT, средство анализа сетевого трафика не может дать достаточно информации для обеспечения мер по защите сетевых ресурсов.

Задача состоит в выделении сетевых адресов и границ полей заголовков протоколов.

Разделение входного массива сетевого трафика на отдельные поля и идентификация сетевых адресов при отсутствии априорных знаний о протоколах является актуальной задачей. При этом предполагаются следующие предположения-эвристики:

– в каждом сетевом пакете присутствует адресная и семантическая части данных;

– адресная часть всегда расположена ближе к началу пакета, чем семантическая;

– адресная часть всегда содержит адреса отправителя и получателя;

– адресная часть меняется реже, чем семантическая.

Известные решения [1-7] предлагают варианты решения частных проблем (унифицированное описание сети, выделение полей из неизвестного трафика одного протокола, классификация трафика на протоколы), но не дают возможности выполнять все действия одновременно.

В [8] представлен способ использования информационной энтропии в качестве метода определения границ полей, позволяющий, используя сравнительно небольшие вычислительные ресурсы, по графикам изменения энтропии отдельных байтов и их взаимной информации делать предположения о структуре анализируемого сетевого протокола. Недостатком такого метода является невозможность его использования на массиве трафика, имеющем более одного протокола.

Предложенный статистико-энтропийный метод, применяет энтропийный модуль для определения границ полей протокола с помощью информационной энтропии и статистический модуль для выделения сетевых адресов на основе анализа статистики вхождения частей пакета в массив трафика.

Статистико-энтропийный метод и его реализация

Для решения проблемы одновременного выделения сетевых адресов и границ семантических полей предлагается объединить два известных алгоритма – статистический и энтропийный (рис. 1). Статистический алгоритм использует оценку количества вхождений похожих на части сетевого пакета подстрок в ранее полученный массив сетевого трафика для выделения из сетевого трафика уровней адресации и конкретных адресов сетевых узлов, а энтропийный с помощью вычисления информационных характеристик осуществляет поддержку решения статистического и определяет границы полей в семантической части.

Входные данные для статистико-энтропийного алгоритма – набор из *lp* сетевых пакетов. Каждый сетевой пакет имеет номер *n* и содержит lb_n байт *d*. Пакет – набор байт $D = (d_i)_{i=1}^{lb_n}$, d_i – байт пакета, расположенный по смещению i от его начала, n – порядковый номер пакета. Набор сетевых пакетов определяется как $DS = (D_i)_{i=1}^{lp}$.

Выходные данные алгоритма – полученный из энтропийного алгоритма набор полей $F = (f_i)_{i=1}^{lf}$, где *lf* – количество выделенных полей, и сформированные из статистического алгоритма множества адресных $AS = (A_i)_{i=1}^{lp}$

и семантических $SS = (S_i)_i^{lp}$ частей сетевых пакетов и сетевых адресов $AddrS = (addr_i)_{i=1}^{lp}$.

Рис. 1. Статистико-энтропийный алгоритм

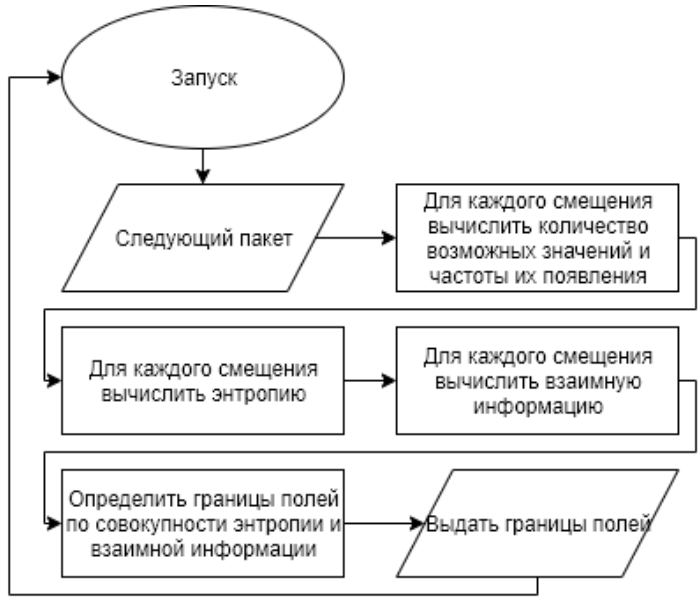


Рис. 2. Энтропийный алгоритм

На основе предложенного в статье [8] метода был разработан новый алгоритм (рис. 2) и предложена программная реализация.

В отличие от алгоритма [8], основанного на раздельном принятии решений, реализовано совместное использование энтропии и взаимной информации для принятия решения о границах полей вместо предложенного раздельного принятия решений.

Для каждого значения *i* оценивается количество вхождений $vh_{i,j}$ каждого значения байта d_i и количество вхождений $vmi_{i,k}$ пар значений. Всего вариантов одиночных значений может быть $2^8 = 256$, поэтому $j \in \overline{(1,256)}$, а пар значений – $(2^8)^2 = 65536$, поэтому $k \in \overline{(1,65536)}$. Красным прямоугольником (рис. 3) показан выбор отдельных значений, зеленым – пары значений с использованием метода скользящего окна.

По формуле $Ph_{ij} = \frac{vh_{i,j}}{lp}$, где i – смещение байта от начала пакета, а j – значение $j \in \overline{(1,256)}$, определяется вероятность появления определённого значения в этой позиции для каждого байта пакета, то есть $i \in \overline{(1,lb)}$, $j \in \overline{(1,256)}$. Аналогично вычисляется $Pmi_{i,j} = \frac{vmi_{i,j}}{lp}$, $i \in \overline{(1,lb_i)}$, $j \in \overline{(1,65536)}$.

По формуле Шеннона рассчитывается 256-чная энтропия для каждого смещения $H_j = -\sum_{i=1}^{lp} Ph_{i,j} \log_{256} Ph_{i,j}$, где j определена

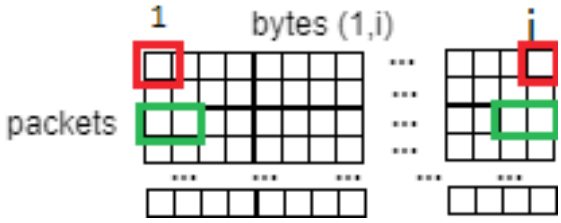


Рис. 3. Результат работы энтропийного алгоритма

так же, как в расчёте вероятности выше. В результате вычислений формируется матрица-строка *H* длиной *max (lb)*.

После расчёта энтропии для каждого смещения от минимального до максимального $j \in \overline{(1, \max(lb)-1)}$ рассчитывается взаимная информация $MI_j = -\sum_{i=1}^{lp} Pmi_{i,j} \log_{65536} \frac{Pmi_{i,j}}{Ph_{i,j} * Ph_{i,j+1}}$.

После того, как энтропия каждого байта и их взаимная информация рассчитаны, можно приступать к определению границ пакета. Создаются матрицы-строки *Hr, Mlr* и *Res* длиной $\max(lb)$, причём Hr_1='start', *Mlr1*='start', *Res1*='start'.

Для $j \in \overline{(2, \max(lb))}$ Hr_j принимает значение *'start'*, если Hr_{j-1}='end', иначе *'field'*, если $H_{j-1}<H_j$, иначе *'end'*. Для $j \in \overline{(2, \max(lb)-1)}$ Mlr_j принимает значение *'start'*, если Mlr_{j-1}='end', иначе *'field'*, если $Ml_{j-1}>Ml_j$, иначе *'end'*.

После определения границ полей пакета для энтропии и взаимной информации в отдельности для принятия совместного решения о расположении границ во входных данных задаётся порог расходимости *T*, который обозначает возможное отличие принятия решения по энтропии и по взаимной информации.

Для каждого $j \in \overline{(2, \max(lb))}$, если $Hr_j=Mlr_j$, то $Res_j=Hr_j$. Иначе если $\exists Hr_{k_1} = 'start'$, $\exists MIr_{k_2} = 'start'$, $|k_1 - j| < T$ и $|k_2 - j| < T$, то $Res_j = Hr_{k_1}$, $Hr_{k_1} = 'field'$, $Hr_j = 'start'$, $MIr_{k_2} = 'field'$, $MIr_j = 'start'$. Для значения $'end'$ аналогично.

Очищается набор полей $F = \emptyset$. Для каждого $j \in \overline{(2, \max(lb))}$, $Res_j = 'start'$, ищется минимальное *k*, такое что *k>j*, и ко множеству *F* добавляется элемент $f_i = \{j, k\}$, имеющий смысл координат начала и конца поля.

При получении на вход нового пакета D_{n+1} для каждого значения *i* обновляется количество вхождений каждого возможного значения байта *di*. Затем по описанному выше алгоритму определяются границы полей, и выполняется переход к ожиданию следующего пакета.

В результате *F* содержит набор границ полей. Их можно использовать для разбора (и ускорения разбора) протокола после того, как будет выделена адресная часть.

Цель статистического алгоритма – выделить из входного потока сетевого трафика адресную и семантическую части, из адресной части – адреса отправителя и получателя.

Перед началом работы алгоритма задаются константы и начальные значения. В качестве констант задаются отношение частот вхождения адресной части к семантической *ADDR_TO_DATA_MULTIPLIER = 5*, значение относительного расстояния Хэмминга, при котором строки считаются похожими *HAMMING_MEASURE_OF_SIMILARITY = 0.1*, в качестве начального значения – средняя длина адресной части *average_border = 1*.

Обработка происходит итеративно по пакетам. Для *i* пакета решения принимаются на основе *i-1* пакетов, обработанных до него.

В целом алгоритм (рис. 4) выглядит так: для каждого пакета D определяются адресная $addr\_part = (d_i)_{i=1}^{la}$ и семантическая $data\_part = (d_i)_{i=la}^{lb}$ части, где *la* – длина адресной части, а *lb* – длина сетевого пакета. В соответствии с полученным значением *la* обновляется средняя длина адресной части *average_border*. Адресная часть $addr\_part$ снова обрабатывается алгоритмом так, будто это весь сетевой пакет *D*, и в этот раз на выходе алгоритма появляются две адресные части верхнего *max_level* и следующего за ним *max_level-1* уровня. Адресная часть верхнего уровня разделяется на две части, каждая из которых добавляется в список сетевых адресов *AS*. Адресная часть уровня *max_level-1* снова обрабатывается алгоритмом для выделения следующих уровней адресации и конкретных сетевых адресов. Так продолжается до тех пор, пока не будет разобран на части сетевой адрес самого низкого уровня.

Полученные адреса позже могут использоваться для построения карты сети.

В общем алгоритме была указана операция выделения адресной и семантической частей, а затем – адресов из адресной части. Для её реализации используется нижеописанный алгоритм (рис. 5).

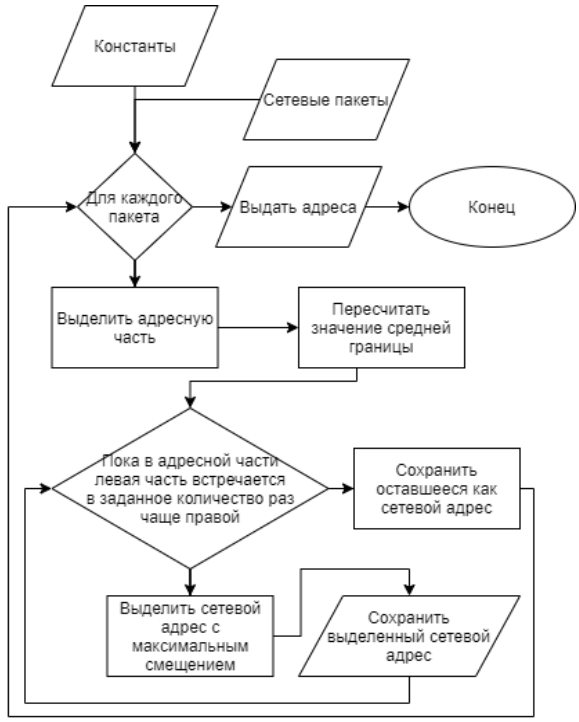


Рис. 4. Общий вид статистического алгоритма

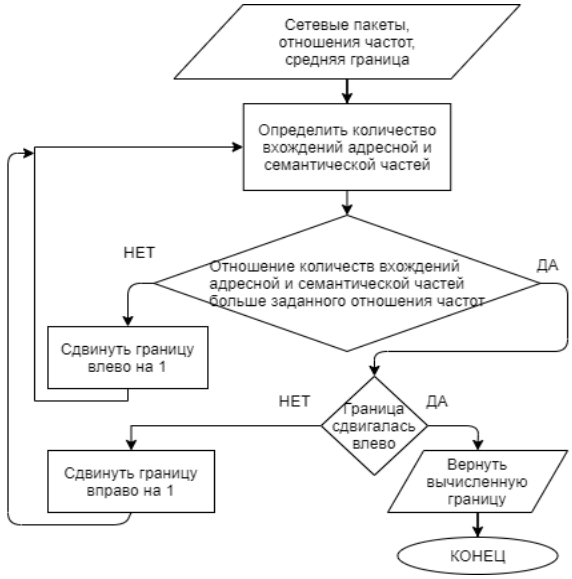


Рис. 5. Определение границы адресной части

Задаётся смещение *border = avarage_ border*, вычисляется количество вхождений $(d_i)_{i=1}^{border}$ и $(d_i)_{i=border}^{lp}$ в обработанный ранее массив трафика. Если количество вхождений первого в *ADDR_TO_DATA_MULTIPLIER* раз больше, чем второго, то *border* увеличивается на 1 и расчёт повторяется. В противном случае действия зависят от того, был ли уже выполнен расчёт для значения *border-1*. Если расчёта ещё не было, то выполняется итеративное уменьшение значения *border* на 1 и перерасчёт количества вхождений $(d_i)_{i=1}^{border}$ и $(d_i)_{i=border}^{lp}$ до тех пор, пока первое не станет встречаться в *ADDR_TO_DATA_MULTIPLIER* раз чаще второго. Если расчёт уже был выполнен, то значение *border* уменьшается на 1.

Если выполнялось выделение адресной и семантической частей из пакета, то по значению смещения border обновляется значение average_border, $(d_i)_{i=1}^{border}$ добавляется в AS, а $(d_i)_{i=border}^{lp}$ – в SS. В противном случае $(d_i)_{i=border}^{lp}$ сохраняется в $AddrS$, а $(d_i)_{i=1}^{border}$ обрабатывается повторно.

Под «количеством вхождений» в описа-

нии алгоритма выделения частей из трафика понимается количество вхождений похожих подстрок в массив строк, полученный из сетевых пакетов. Из-за особенностей сетевого трафика (поля-разделители, поля-идентификаторы) принято решение искать не абсолютно совпадающие строки, а похожие, причём в качестве алгоритма оценки похожести использовать модифицированный метод Хэмминга.

Модификация состоит в том, что расстояние Хэмминга считается не абсолютное, а относительно длины строки $hamming_{sim} = \frac{dist}{num\_symbols}$, где *dist* – расстояние Хэмминга, *num_symbols* – количество символов в сравниваемых подстроках. Таким образом, вместо термина «расстояние» уместнее использовать «коэффициент схожести».

Для оценки количества вхождений строка *D* по предполагаемой длине адресной части делится на адресную $(d_i)_{i=1}^{border}$ и семантическую $(d_i)_{i=border}^{lp}$ части. Для каждой части оценивается коэффициент схожести *sim_coef* с остальными обработанными пакетами. В случае, если *sim_coef* оказывается больше *HAMMING_MEASURE_OF_SIMILARITY*, то количество оцениваемой подстроки вхождений увеличивается.

Такой метод оценки количества вхождений подстрок вычислительно затратен. Поэтому предлагается альтернативный метод: строить дерево всех возможных комбинаций, и в листьях хранить количество появлений каждой из них.

Сетевой пакет представляет собой набор байт, каждый из которых может принимать 256 значений. Учитывая ограниченность реальных вычислительных ресурсов, предлагается использовать последовательности длиной 4 байта, в этом случае, с одной стороны, 8ГБ оперативной памяти достаточно для работы, и, с другой, последовательности не будут слишком короткими.

В таком случае скорость поиска количества вхождений заданной комбинации будет линейно зависеть от её длины, что даст возможность увеличивать размер окна при анализе больших объёмов трафика.

В результате обработки набора сетевых пакетов двумя разработанными алгоритмами имеются набор полей $F = (f_i)_{i=1}^{lf}$, множества адресных $AS = (A_i)_{i=1}^{lp}$ и семантических $SS = (S_i)_{i=1}^{lp}$ частей сетевых пакетов и дерево сетевых адресов $AddrS = (addr_i)_{i=1}^{lp}$.

Семантические части $SS = (S_i)_{i=1}^{lp}$ пакетов делятся на сегменты по известному набору полей *F*, формируя множество наборов семантических полей $FS = ((fs_i)_{i=1}^{lp})_{j=1}^{lf}$. Затем *FS* и *AddrS* передаются во внешнюю систему для построения графической топологии сети и отображения типов узлов на основе *FS* для каждого узла.

Анализ трафика, содержащего более одного протокола

Используемый энтропийный алгоритм позволяет быстро обрабатывать трафик, в состав которого входят пакеты, сформированные исключительно по одному протоколу (или одной иерархической группе протоколов). Для обхода этого ограничения предлагается для входного трафика строить на основе длин энтропийных полей дерево протоколов, и на каждом сетевом уровне анализировать их раздельно (рис. 6).

Дерево протоколов представляет собой иерархический граф, корнем которого является родительский протокол (протокол самого нижнего уровня, например, Ethernet), а узлами – протоколы следующих уровней, причём кратчайшее расстояние от корня до узла определяет уровень узла.

Задаётся структура *Node*, содержащая порядковый номер родителя на предыдущем уровне *parent_id*, длину первого поля текущего узла *field_len*, номера пакетов, относящиеся к текущему узлу *num_packets*, суммарную длину первых полей до начала пакета *sum_len*.

Создаётся список уровней *nodes*, каждый из которых является списком узлов (структур *Node*). Таким образом, *nodes[level][id_on_level]* однозначно определят узел, где *level* – порядковый номер уровня в дереве протоколов, а *id_on_level* – порядковый номер ветви на уровне.

В списке структур *nodes* создаётся нулевой уровень с единственным узлом – корневым протоколом. Для него с помощью энтропийного алгоритма выбирается длина первого поля *field_len*, и из трафика выбираются все уникальные значения этого поля. Для каждого уникального значения на следующем уровне списка структур *nodes* создается новый узел.

Исходя из предположения о том, что адресная часть всегда находится ближе к началу пакета, чем часть данных, выбирается глубина анализа (количество байт от начала пакета, которое будет рассматриваться). Анализ продолжается до тех пор, пока не будет достигнут этот предел.

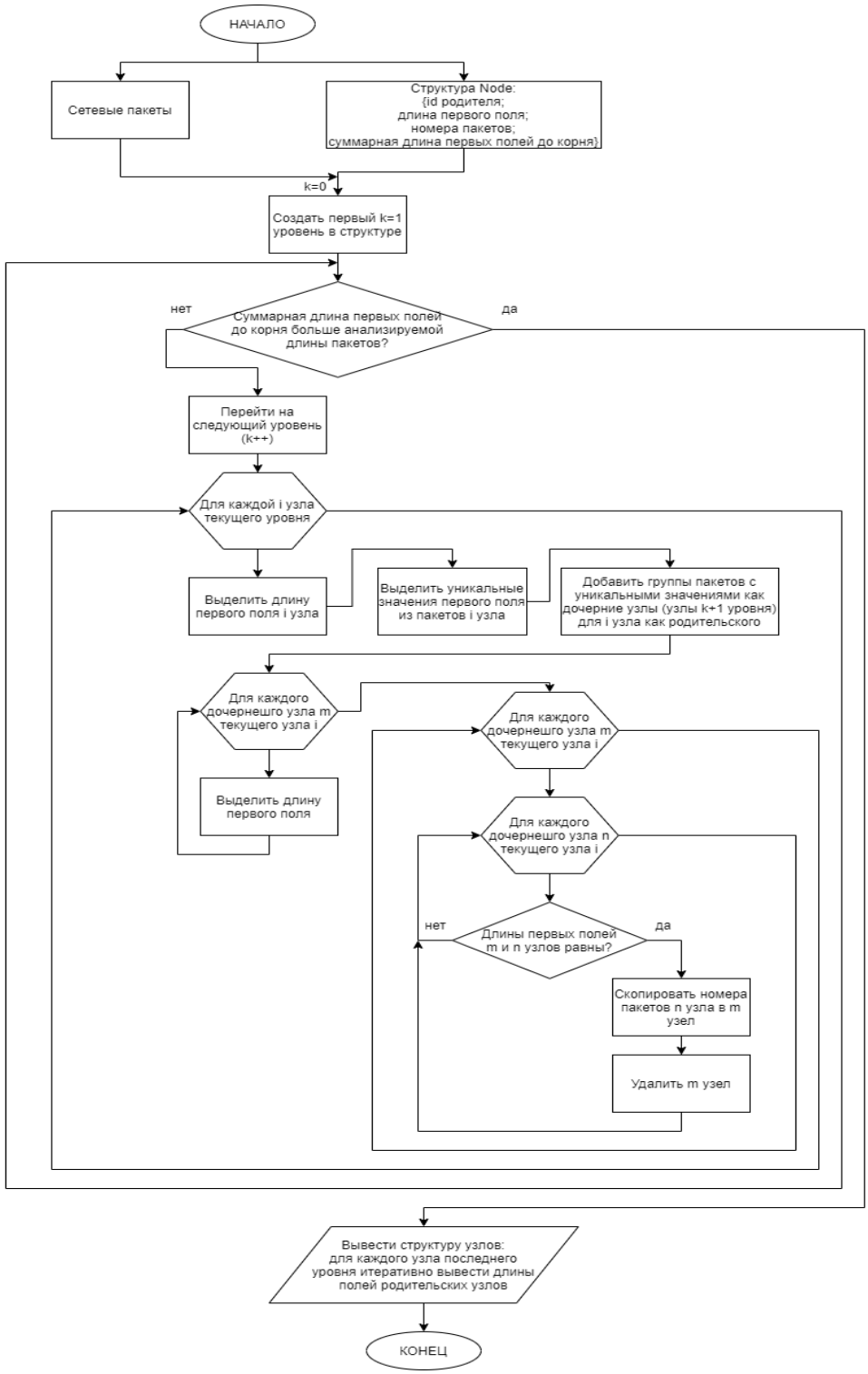


Рис. 6. Итеративный энтропийный алгоритм

Выполняется переход на следующий уровень структуры *nodes,* на котором уже существуют несколько узлов, соответствующих уникальным значениям первого поля родительского узла. Для каждого из узлов с помощью энтропийного алгоритма выбирается длина первого поля, из принадлежащих узлу пакетов выбираются все уникальные значения этого первого поля, на следующем уровне для каждого из них создаются узлы, значение *parent_id* которых соответствует порядковому номеру родителя (текущего узла), а в

список пакетов *num_packets* записываются номера таких пакетов родительского узла, первые несколько байт которых соответствуют выбранному уникальному значению. Затем для каждого из созданных узлов следующего уровня с помощью энтропийного алгоритма выбирается длина первого поля, и те узлы, у которых его значения совпали, объединяются в один узел.

После достижения заданного количества байт от начала пакета анализ завершается, и формируется дерево протоколов. При этом узлы, расположенные ближе к корневому, описывают более низкоуровневые протоколы, чем узлы, расположенные дальше от корневого.

Результаты работы алгоритмов

Программная реализация статистического метода позволяет получить массивы длин адресных частей и предполагаемые адреса. При наложении полученных данных на реальный сетевой трафик видно (рис. 7), что верно определяется длина заголовка Ethernet и входящие в него MAC-адреса (выделено зелёным), но при этом к адресу источника ошибочно добавляются служебное редко меняющееся поле type (выделено красным).

```
>>> print(src_addr[7], dst_addr[7])
  b'\xb4\xb5/t\x08\x9d' b'\xac\xc1\xd6b\xc2\x08\x00'
Offset(h)  00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F
000003E0   9B A5 AD 50 11 7E 12 B1 EE 00 00 8A 18 7A 5D 18
000003F0   A6 00 00 3C 00 00 00 3C 00 00 00 B4 B5 2F 74 08
00000400   9D F4 AC C1 D6 62 C2 08 00 45 00 00 28 04 03 40
00000410   00 7E 06 46 F9 C0 A8 0C 75 C0 A8 24 0E 32 C8 24
00000420   47 31 9B A5 AD 31 BF E0 69 50 10 80 00 3D 7F 00
00000430   00 00 00 00 00 00 00 8A 18 7A 5D 50 A7 00 00 3C
```

Рис. 7. Результат работы скрипта (статистический алгоритм)

В результате обработки сетевого трафика с помощью программы, реализующей энтропийный метод, построены графики энтропии и взаимной информации.

Анализируя график энтропии (рис. 8) по модели, описанной выше, можно видеть, что энтропия возрастает (или хотя бы не убывает) на длине всего поля и уменьшается в его конце. По этому графику можно с некоторой вероятностью выделить MAC-адреса, подтверждая полученные статистическим алгоритмом результаты.

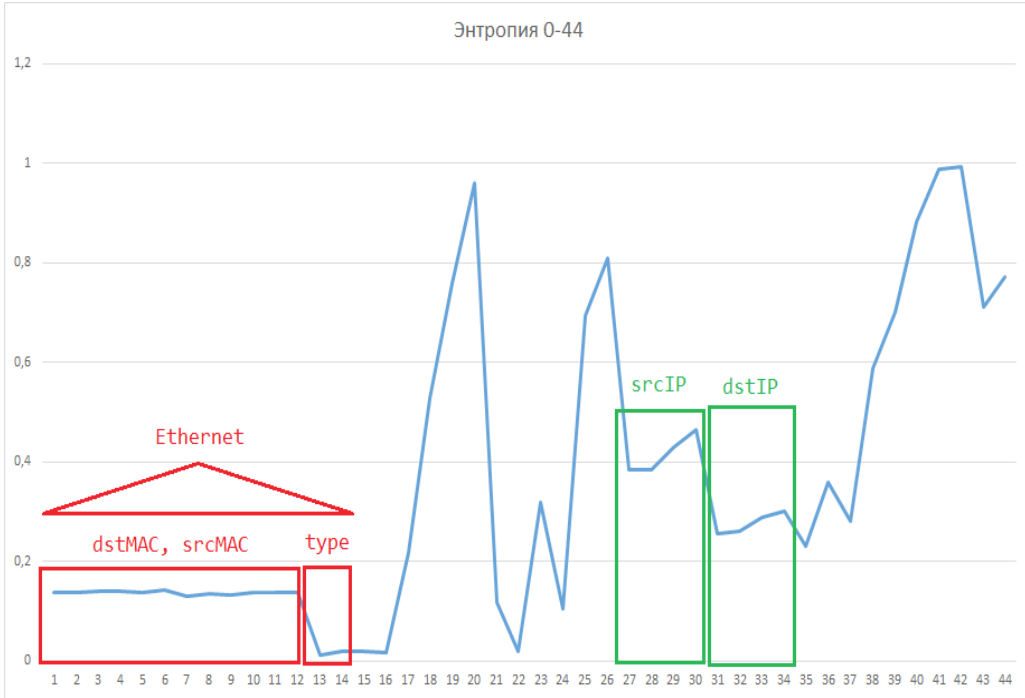


Рис. 8. Энтропия

График взаимной информации (рис. 9) показывает уменьшение значения в конце полей, описывающих IP-адреса источника и получателя, но в целом не поддаётся описанию, поэтому для разделения полей предлагается использовать энтропию и взаимную информацию совместно, принимая решение так, как описано в модели энтропийного алгоритма.

Объединение результатов в рамках статистико-энтропийного метода позволяет наложением полей из энтропийного алгоритма подтвердить полученную статистическим методом информацию о MAC-адресах. Дополнительная информация о разделении на поля может быть использована для классификации обнаруженных узлов сети.

Итеративный энтропийный алгоритм (Рисунок 6) в результате работы позволяет определить длины использованных протоколов даже в случае, если их было несколько.

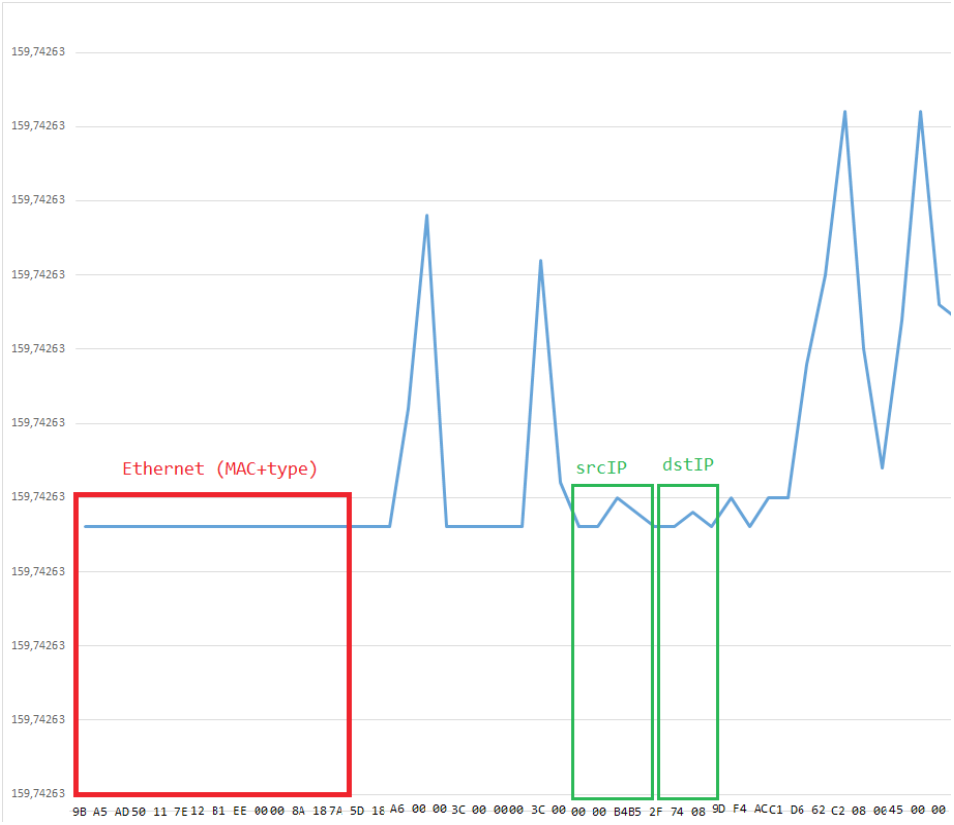


Рис. 9. Взаимная информация

Выводы

Разработанные алгоритмы описывают метод анализа сетевого трафика при отсутствии априорных сведений о нём. Программная реализация решает поставленную задачу выделения из трафика адресов (статистический) и разделения трафика на поля (энтропийный), а их объединение позволяет получать информацию не только о наличии или отсутствии сетевых узлов, но и давать им характеристику. Разработанный итеративный энтропийный метод решает задачу анализа трафика, содержащего пакеты, относящиеся к различным протоколам.

Дальнейшее развитие проекта будет заключаться в оптимизации коэффициентов, что позволит получать более точные результаты, и в доработке итеративного энтропийного алгоритма для увеличения его производительности.

Литература / References

1. Dmitri Bekerman, Bracha Shapira, Lior Rokach, Ariel Bar “Unknown Malware Detection Using Network Traffic Classification”, 2015 IEEE Conference on Communications and Network Security (CNS), Florence, Italy, 28-30 Sept. 2015, pp. 134-142, DOI: 10.1109/CNS.2015.7346821.

2. Rui Li, Xi Xiao, Shiguang Ni, Haitao Zheng, Shutao Xia “Byte Segment Neural Network for Network

Traffic Classification", 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS), Banff, AB, Canada, 4-6 June 2018, DOI: 10.1109/IWQoS.2018.8624128.

3. Antônio J.Pinheiro, Jeandrode M. Bezerra, Caio A.P.Burgardt, Divanilson R.Campelo "Identifying IoT devices and events based on packet length from encrypted traffic", Computer Communications, Volume 144, 15 August 2019, Pages 8-17, DOI: 10.1016/j.comcom.2019.05.012.

Аветисян А.И. , Гетьман А.И. Восстановление структуры бинарных данных по трассам программ. Труды Института системного программирования РАН, том 22, 2012, с. 95-118. DOI: 10.15514/ISPRAS-2012-22-7. [Avetisyan A.I., Get'man A.I. Vosstanovleniye struktury binarnykh dannykh po trassam programm. Trudy Instituta sistemnogo programmirovaniya RAN, tom 22, 2012, s. 95-118. DOI: 10.15514/ISPRAS-2012-22-7].

4. Shaun Voigt, Catherine Howard, Dean Philp and Christopher Penny "Representing and Reasoning about Logical Network Topologies" In book: Graph Structures for Knowledge Representation and Reasoning, 2018, DOI: 10.1007/978-3-319-78102-0_4.

5. Weidong Cui, Jayanthkumar Kannan, Helen J. Wang "Discoverer: Automatic protocol reverse engineering from network traces" Proceedings of 16th USENIX Security Symposium, 6-10 August 2007 Article No.: 14 Pages 1–14.

6. João Antunes Nuno Neves Paulo Verissimo "Reverse Engineering of Protocols from Network Traces", 2011 18th Working Conference on Reverse Engineering, Limerick, Ireland, 17-20 Oct. 2011, DOI: 10.1109/WCRE.2011.28.

7. Fanghui Sun Shen Wang, Chunrui Zhang, Hongli Zhang "Unsupervised field segmentation of unknown protocol messages", Computer Communications, Volume 146, 15 October 2019, Pages 121-130, DOI: 10.1016/j.comcom.2019.06.013.

ДОМУХОВСКИЙ Николай Анатольевич, заместитель генерального директора по научно-технической работе ООО «Уральский центр систем безопасности» (ООО «УЦСБ»). 620100, г. Екатеринбург, ул. Ткачей, 23. E-mail: ndomukhovsky@ussc.ru

СИНАДСКИЙ Алексей Николаевич, младший инженер ООО «Уральский центр систем безопасности» (ООО «УЦСБ»). 620100, г. Екатеринбург, ул. Ткачей, 23. E-mail: alexsin@e1.ru

DOMUKHOVSKY Nikolay, Deputy General Director for Scientific and Technical Work LLC Ural Center for Security Systems (LLC USSC). 620100, Yekaterinburg, Tkachey str., 23. E-mail: ndomukhovsky@ussc.ru

SINADSKIY Alexey, junior engineer LLC Ural Center for Security Systems (LLS USSC). 620100, Yekaterinburg, Tkachey str., 23; e-mail: alexsin@e1.ru

УДК 004.056.5:378.016 Вестник УрФО № 1(35) / 2020, с. 58–64

Данилов А. Н., Шабуров А. С., Южаков А. А.

DOI: 10.14529/secur200106

О СОВЕРШЕНСТВОВАНИИ ПОДГОТОВКИ СПЕЦИАЛИСТОВ ПО ЗАЩИТЕ ИНФОРМАЦИИ В ПЕРМСКОМ НАЦИОНАЛЬНОМ ИССЛЕДОВАТЕЛЬСКОМ ПОЛИТЕХНИЧЕСКОМ УНИВЕРСИТЕТЕ (ПНИПУ). К 60 - ЛЕТНЕМУ ЮБИЛЕЮ КАФЕДРЫ «АВТОМАТИКА И ТЕЛЕМЕХАНИКА»

В статье анализируется проблема подготовки специалистов по защите информации, современные актуальные аспекты, формирующие образ будущего выпускника. Приводится историческая справка, фактические данные, иллюстрирующие процесс развития системы подготовки специалистов по защите информации на кафедре автоматики и телемеханики Пермского национального исследовательского политехнического университета (ПНИПУ). Рассматриваются основные итоги образовательной деятельности кафедры в подготовке специалистов по защите информации, устанавливаются основные ориентиры и перспективы дальнейшего развития. Приводится модель иерархической структуры формирования элементов профессиональной компетенции выпускника.

Ключевые слова: *информационная безопасность, защита информации профессиональный стандарт, компетентностная модель.*

ON IMPROVING THE SYSTEM OF TRAINING INFORMATION SECURITY SPECIALISTS AT THE PERM NATIONAL RESEARCH POLYTECHNIC UNIVERSITY (PNRPU). TO THE 60TH ANNIVERSARY OF THE DEPARTMENT «AUTOMATION AND TELEMECHANICS»

The article analyzes the problem of training information security specialists, current relevant aspects that shape the image of a future graduate. Providing historical background, factual data illustrating the process of developing a system for training information security specialists at the Department of Automation and Telemechanics of Perm National Research Polytechnic University (PNRPU). The main educational activities result of the department in the preparation of information security specialists are considered, the main guidelines and prospects for further development are established. A hierarchical structure model of the formation elements of graduate professional competence is given.

Keywords: *information security, information protection, professional standard, competency model.*

Совершенствование качества подготовки специалистов по защите информации является одним из основополагающих ориентиров в деятельности учебного заведения, реализующего образовательные программы УГСНП 10.00.00 Информационная безопасность. При этом, ответственность за подготовленность выпускников определяется, как на уровне законодательства Российской Федерации [1], так и требованиями перехода экономической модели развития страны в область цифровых технологий [2]. Обеспечить потребности рынка труда в квалифицированных специалистах необходимо в кротчайшие сроки, с учетом государственных приоритетов развития страны, тенденций развития информационных систем, специфики деятельности предприятий и организаций, а также региональных аспектов обеспечения информационной безопасности.

В данных условиях немаловажное значение приобретает ранее приобретенный опыт и сложившиеся традиции конкретной образовательной организации, базовые принципы обучения, учебно-лабораторная база, позволяющая формировать основы фундамен-

тальных, инженерных, специальных знаний и умений будущих специалистов. В то же время, необходим постоянный поиск и внедрение инновационных решений, методических приемов и форм занятий, позволяющих в установленные сроки сформировать необходимые профессиональные компетенции выпускника.

Одним из примеров успешной реализации нескольких направлений обучения в рамках одного структурного подразделения, является кафедра «Автоматика и телемеханика» (АТ) электротехнического факультета (ЭТФ) Пермского национального исследовательского политехнического университета (ПНИПУ). Образовательная деятельность данной кафедры базируется на сформировавшихся за десятилетия традициях инженерной школы, опыте квалифицированных педагогов-наставников, инновационных подходах к обучению, одновременно по нескольким направлениям подготовки.

Кафедра АТ была организована через несколько лет после образования университета (в то время Молотовского горного института) – в 1960 году. Необходимость её открытия была обусловлена быстро растущими потребностями предприятий промышленности Перми и всего региона в специалистах в области автоматизации технологических процессов и производств. Поэтому кафедра АТ положила начало созданию всего ЭТФ ПНИПУ, на сегодня состоящего уже из пяти выпускающих кафедр [3].

Выпускники кафедры АТ по специальности «Автоматика и телемеханика» (после переименования «Управление и информатика в технических системах») всегда были активно востребованы на предприятиях Пермского региона.

Для решения задач надёжной передачи информации в проектируемых и реализуемых системах, прежде всего, предприятий добывающих и перерабатывающих отраслей (ЛУКойл, Газпром), разработке и производстве радиоэлектронных устройств (промышленные предприятия «Морион», «Такт», «Пермская научно-производственная приборостроительная компания», «Телта»), организации и эксплуатации вычислительных и телекоммуникационных сетей («Ростелеком», операторы мобильной связи) стала очевидной необходимость открытия на базе кафедры АТ специальности «Телекоммуникации». Данное направление было открыто в 1998 году, при этом специализацией была выбрана наиболее востребованная для региона – «Сети связи и системы коммутации».

Начиная с 2004 г., на основе сложившегося опыта подготовки инженеров по управлению в технических системах и в сфере телекоммуникации и связи на кафедре АТ была организована подготовка специалистов по защите информации. На первоначальном этапе становления специальности были решены основные задачи, связанные с формированием учебно-лабораторной базы, методическим обеспечением учебного процесса. В 2007 г., в соответствии с приказом Федерального агентства по образованию на кафедре организован Региональный учебно-научный центр (РУНЦ) по защите информации. Благодаря данному решению, помимо реализации основных образовательных программ высшего образования, началась работа по повышению квалификации специалистов.

За период с 2009 по 2018 г только по программе «Комплексное обеспечение информационной безопасности автоматизированных систем» на базе РУНЦ прошли повышение квалификации более 700 сотрудников территориальных учреждений Банка России. Данный вклад в переподготовку специалистов по защите информации был отмечен в 2013 г. дипломом Национального форума по информационной безопасности «ИНФОФОРУМ», в номинации «Образовательный центр года».

Успешной реализации программ подготовки кадров способствовало взаимовыгодное сотрудничество с органами исполнительной власти: Управлением Роскомнадзора по Пермскому краю, Радиочастотной службой Приволжского федерального округа, предприятиями и организациями г. Перми: АО «ОДК-СТАР», АО «ОДК-Авиадвигатель», АО «Гознак» ООО «Лукойл-Информ», ЗАО «ИвсСети», ЗАО «БИОНТ», ЗАО «Проминформ» и др. Взаимодействие с данными организациями и предприятиями позволило осуществлять качественную подготовку специалистов по защите информации, вести научные исследования по актуальным проблемам информационной безопасности.

Основными направлениями научно-исследовательских работ кафедры в области информационной безопасности являются разработка информационных систем в защищенном исполнении, исследования методов комплексной защиты объектов информатиза-

ции, защита информационных систем от кибератак, управление информационной безопасностью, модели и методы распознавания образов в информационных системах.

Наибольший вклад в становление и развитие направления подготовки по информационной безопасности внесли ученые и преподаватели кафедры: д.т.н., профессор Матушкин Н.Н., д.т.н., профессор Южаков А.А., к.т.н., доцент Данилов А.Н., к.т.н., доцент Шабуров А.С, к.т.н. доцент Безукладников И.И., к.т.н. Полшков А.В., к.т.н. Кокоулин А.Н. Всего, за годы обучения по направлению «Информационная безопасность» было подготовлено более 500 бакалавров и специалистов по защите информации, большинство из которых трудятся на предприятиях и в организациях г. Перми и Пермского края.

В настоящее время кафедра АТ осуществляет подготовку по образовательным программам направления «Информационная безопасность» на уровне бакалавриата и магистратуры, а также по специальности «Информационная безопасность автоматизированных систем». Ведется плановая, систематическая работа по совершенствованию учебно-лабораторной и методической базы, направленная на повышение качества подготовки выпускников. Для проведения занятий привлекаются ведущие специалисты по защите информации. Например, в течение многих лет для проведения занятий по дисциплине «Программно-аппаратные средства защиты информации» привлекается руководитель направления по созданию и эксплуатации защищенных информационных систем отдела информационной безопасности АО «Гознак» к.т.н. Капгер И.В (г. Москва).

Качественный рост уровня подготовки профессорско-преподавательского состава, студентов кафедры позволил продемонстрировать некоторые из достижений. В ходе Методического сбора 12 апреля 2018 г., проводившегося Управлением ФСТЭК по Приволжскому федеральному округу, на базе кафедры были успешно организованы и проведены показные занятия с представителями органов власти и специалистами по технической защите информации. Созданные на базе кафедры лабораторные стенды по защите информации, а также применение Интерактивного учебно-лабораторного комплекса по информационной безопасности «СОТСБИ-Guard» позволяют отрабатывать различные учебные задачи, направленные на блокирование утечек конфиденциальной информации, противодействие компьютерным атакам, анализ уязвимости информационных систем и т.п.

Внедрение новых образовательных технологий, методов и форм интерактивного обучения, а также заинтересованность предприятий в подготовке квалифицированных специалистов по защите информации заставляют искать новые и совершенствовать имеющиеся образовательные приемы и методы. Благодаря студенческой инициативе и усилиям преподавательского состава кафедры, в 2019 году на базе ПНИПУ была развернута необходимая инфраструктура для проведения всероссийских соревнований в формате CTF.

Созданные на кафедре условия для творческого роста позволяют молодым и талантливым преподавателям развиваться во многих перспективных проектах, решать исследовательские задачи на стыке научных направлений. В 2020 году, в рамках Всероссийского конкурса молодых специалистов и образовательных центров в области информационной безопасности «ИНФОФОРУМ - НОВОЕ ПОКОЛЕНИЕ» дипломом лауреата в номинации «Молодой специалист года» был награжден к.т.н., доцент кафедры Каменских А.Н.

В то же время, предстоит решить еще немало практических и научно-исследовательских задач в поиске наиболее эффективных приемов и методов подготовки квалифицированных кадров, в том числе и специалистов по защите информации. Одним из актуальных научно-исследовательских направлений кафедры является разработка инновационных учебно-лабораторных комплексов и стендов, поиск эффективных моделей и методов обучения, направленных на формирование требуемых профессиональных компетенций выпускников.

Как правило, проблема формирования компетенций решается через реализацию у студента соответствующей компонентной структуры (знаний, умений, владений), а также навыков и опыта их практического применения. Каждый компонент, в свою очередь, состоит из набора элементов. При их формировании применимы традиционные образовательные технологии, например, лекции, семинары, лабораторные практикумы, курсовое проектирование и т.д. Основная задача заключается в подборе для формирования элементов компетенции таких видов работы студентов, чтобы обеспечить требуемый уровень ее освоения [4].

Для оценки уровня освоения компетенций необходимо выбрать средства и методы контроля элементов компетенций. Данная задача связана с формами контроля, при помощи которых можно проверить, сформировались ли в процессе работы студентов (аудиторной или самостоятельной) закрепленные за ними элементы и как именно. Это необходимо для проведения текущей аттестации, а также для управления качеством учебного процесса (выявление слабых мест в подготовке, сложных для понимания вопросов и т.п.) [5].

Компонентная структура дисциплинарной компетенции, а также средства формирования и контроля элементов компетенции, с учетом иерархического принципа построения принятой модели, приведены на рис. 1.

В свою очередь, дисциплинарная компетенция, формируемая в процессе освоения дисциплины «Управление информационной безопасностью» детализируется на элементы дисциплинарной компетенции, а именно:

Знать:

$З_1$ – цели и задачи управления информационной безопасностью;

$З_2$ – стандарты систем и процессов управления информационной безопасностью;

$З_3$ – порядок оценки рисков информационной безопасности;

$З_4$ – методы обработки рисков информационной безопасности.

Уметь:

$У_1$ – разрабатывать частные политики информационной безопасности;

$У_2$ – оценивать информационные риски.

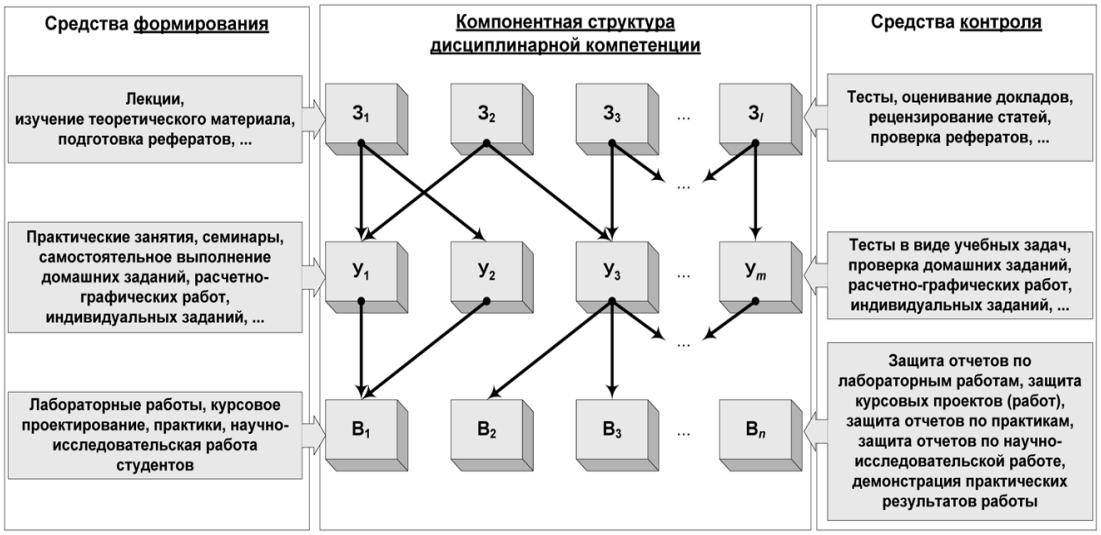


Рис. 1. Модель формирования и контроля компонентной структуры дисциплинарной компетенции

Проиллюстрируем изложенные в данном разделе рекомендации по формированию иерархии тестовых заданий проверки элементов одной профессиональной компетенции, используя иерархическую модель ее представления [6]. Сделаем это на примере формирования тестовых заданий для элементов компетенции, закрепленной за дисциплиной «Управление информационной безопасностью» учебного плана направления подготовки 10.03.01 «Информационная безопасность» (бакалавриат). Профессиональная компетенция – «Способность участвовать в работах по реализации политики информационной безопасности, применять комплексный подход к обеспечению информационной безопасности объекта защиты» является основой для формирования уже дисциплинарной компетенции.

Владеть:

$В_1$ – методами оценки информационных рисков при реализации политики информационной безопасности.

На рис. 2 показана модель иерархической структуры формирования элементов дисциплинарной компетенции для рассматриваемого примера.

Также возможен вариант структуры, когда имеют места разветвления (один элемент участвует в формировании нескольких элементов следующего уровня иерархии). Очевидно, что формулировки элементов могут быть подобраны для реализации различных топологий. При этом выбор модели структуры формирования компетенции может выполнять разработчик рабочей программы учебной дисциплины [7].

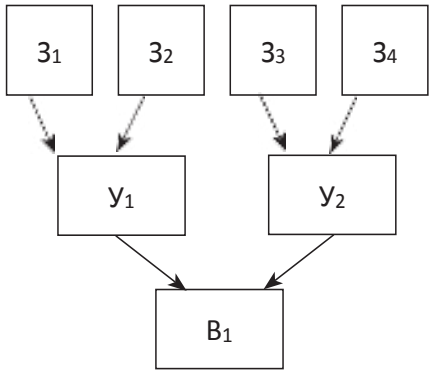


Рис. 2. Граф модели формирования элементов дисциплинарной компетенции

Таким образом, сформировавшаяся на кафедре АТ система образовательной деятельности, базируется на лучшем опыте освоения базовых инженерных специальностей, педагогическом мастерстве профессорско-преподавательского состава, инновационных подходах, что позволяет развиваться и совершенствоваться направлению подготовки по информационной безопасности, а также дает возможность для открытия новых, востребованных направлений и специальностей.

В 2020 году на кафедре организована подготовка по еще одному перспективному образовательному направлению – «Робототехника» на уровне бакалавриата и магистратуры. Это позволит реализовать новые учебные программы и планы, всесторонне развиваться преподавательскому составу, решать востребованные и современные научные проблемы, даст новый импульс для развития кафедры АТ на долгие годы.

Литература

1. Федеральный закон «Об образовании в Российской Федерации» от 29.12.2012 № 273-ФЗ.[Электронный ресурс]: Доступ из справ.-правовой системы «КонсультантПлюс».

2. Мовчан И.Н. Проблемы подготовки специалистов в области информационной безопасности Открытое образование 2013 № 5. С. 78–80.

3. Кон Е. Л., Фрейман В. И., Южаков А. А. К 60-летнему юбилею кафедры автоматики и телемеханики Пермского национального исследовательского политехнического университета // Вестник Поволжского государственного технологического университета. Сер.: Радиотехнические и инфокоммуникационные системы. 2019. № 4 (44). С. 85–91. DOI: 10.25686/2306-2819.2019.4.85

4. Кон Е.Л., Фрейман В.И., Южаков А.А. К вопросу о контроле элементов дисциплинарных компетенций в рамках основной образовательной программы (на примере технических направлений подготовки) // Открытое образование. – 2013. – № 3 – С. 12–19.

5. Кон Е.Л., Фрейман В.И., Южаков А.А. Оценка качества формирования компетенций студентов технических вузов при двухуровневой системе обучения // Научные исследования и их практическое применение. Современное состояние и пути развития '2012: сб. науч. тр. междунар. науч.-практ. конф., 2–12 октября 2012 г. – Одесса: КУПРИЕНКО, 2012. – Т. 9. – С. 39–41.

6. К вопросу о подготовке и оценке компетенций выпускников высшей школы с использованием модулей «Вектор развития направления» и «Квалификационные требования работодателей» / Е.Л. Кон и др. // Открытое образование. – 2012. – № 3 – С. 17–29.

7. Фрейман В.И. Разработка учебно-методического комплекса дисциплины в соответствии с ФГОС нового поколения // Вестник Пермского государственного технического университета. Электроника, информационные технологии, системы управления. – 2009. – № 3. – С. 47–50.

References

1. Federal'nyy zakon «Ob obrazovanii v Rossiyskoy Federatsii» ot 29.12.2012 № 273-FZ.[Elektronnyy resurs]: Dostup iz sprav.-pravovoy sistemy «Konsul'tantPlyus». 2. Movchan I.N. Problemy podgotovki spetsialistov v oblasti informatsionnoy bezopasnosti Otkrytoye obrazovaniye 2013 № 5. S. 78-80.

3. Kon Ye. L., Freyman V. I., Yuzhakov A. A. K 60-letnemu yubileyu kafedry avtomatiki i telemekhaniki Permskogo natsional'nogo issledovatel'skogo politekhnicheskogo universiteta // Vestnik Povolzhskogo

gosudarstvennogo tekhnologicheskogo universiteta. Ser.: Radiotekhnicheskiye i infokommunikatsionnyye sistemy. 2019. № 4 (44). S. 85-91. DOI: 10.25686/2306-2819.2019.4.85

4. Kon Ye.L., Freyman V.I., Yuzhakov A.A. K voprosu o kontrole elementov distsiplinarnykh kompetentsiy v ramkakh osnovnoy obrazovatel'noy programmy (na primere tekhnicheskikh napravleniy podgotovki) // Otkrytoye obrazovaniye. – 2013. – № 3 – S. 12–19.

5. Kon Ye.L., Freyman V.I., Yuzhakov A.A. Otsenka kachestva formirovaniya kompetentsiy studentov tekhnicheskikh vuzov pri dvukhurovnevoy sisteme obucheniya // Nauchnyye issledovaniya i ikh prakticheskoye primeneniye. Sovremennoye sostoyaniye i puti razvitiya '2012: sb. nauch. tr. mezhdunar. nauch.-prakt. konf., 2–12 oktyabrya 2012 g. – Odessa: KUPRIYENKO, 2012. – T. 9. – S. 39–41.

6. K voprosu o podgotovke i otsenke kompetentsiy vypusknikov vysshey shkoly s ispol'zovaniyem moduley «Vektor razvitiya napravleniya» i «Kvalifikatsionnyye trebovaniya rabotodateley» / Ye.L. Kon i dr. // Otkrytoye obrazovaniye. – 2012. – № 3 – S. 17–29.

7. Freyman V.I. Razrabotka uchebno-metodicheskogo kompleksa distsipliny v sootvetstvii s FGOS novogo pokoleniya // Vestnik Permskogo gosudarstvennogo tekhnicheskogo universiteta. Elektronika, informatsionnyye tekhnologii, sistemy upravleniya. – 2009. – № 3. – S. 47–50.

ДАНИЛОВ Александр Николаевич, кандидат технических наук, доцент кафедры автоматики и телемеханики. Пермский национальный исследовательский политехнический университет. 614990, г. Пермь, Комсомольский пр., 29. E-mail: dan@pstu.ru

ШАБУРОВ Андрей Сергеевич, кандидат технических наук, доцент кафедры автоматики и телемеханики. Пермский национальный исследовательский политехнический университет. 614990, г. Пермь, Комсомольский пр., 29. E-mail: shans@at.pstu.ru

ЮЖАКОВ Александр Анатольевич, доктор технических наук, профессор, заведующий кафедрой автоматики и телемеханики. Пермский национальный исследовательский политехнический университет. 614990, г. Пермь, Комсомольский пр., 29. E-mail: uz@at.pstu.ru

DANILOV Alexandr, Candidate of Technical Sciences, Associate Professor of the Department of Automation and Telemechanics. Perm National Research Polytechnic University. 614990, Perm, Komsomolsky Ave, 29. Email: dan@at.pstu.ru

SHABUROV Andrey, Candidate of Technical Sciences, Associate Professor of the Department of Automation and Telemechanics. Perm National Research Polytechnic University. 614990, Perm, Komsomolsky Ave, 29. Email: shans@at.pstu.ru

YUZHAKOV Aleksandr, Doctor of Engineering Sciences, Professor, the Head of the Chair Automatics and Telemechanics, Perm National Research Polytechnic University. 614990, Perm, Komsomolsky Ave, 29. E-mail: uz@at.pstu.ru

УДК 004.056.32 + 343.142.91 Вестник УрФО № 1(35) / 2020, с. 65–71

Минбалеев А. В.

DOI: 10.14529/secur200107

ИСПОЛЬЗОВАНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И ПРАВОВОЕ ОБЕСПЕЧЕНИЕ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ И КИБЕРБЕЗОПАСНОСТИ В РОССИИ И ЗА РУБЕЖОМ: ОСНОВНЫЕ ПРОБЛЕМЫ[1]

Статья посвящена анализу основных проблем обеспечения информационной безопасности в Российской Федерации и в мире в процессе создания и использования искусственного интеллекта, которые являются стратегическим направлением в области обеспечения безопасности личности, общества и государства. Весьма актуальной сегодня становится проблема замещения искусственным интеллектом человека в процессе его собственной жизнедеятельности. Она поднимается сегодня, поскольку человек прекращая совершать значительное количество операций по поиску, обработке, передаче информации становится полностью зависимым от роботов, доверяет им все больше и готов пожертвовать рядом своих свобод, законных интересов, в том числе и в сфере информационной безопасности, в ходе передачи тех или иных видов человеческих процессов роботам.

В статье анализируются проблемы обеспечения информационной безопасности при использовании искусственного интеллекта при осуществлении террористической деятельности, а также использования искусственного интеллекта в противодействии терроризма. Анализируются угрозы информационной безопасности при осуществлении массового наблюдения и анализа данных с использованием технологий искусственного интеллекта, в том числе в процессе противодействия преступности.

Ключевые слова: *информационная безопасность, искусственный интеллект, противодействие преступности, массовой наблюдение, вызовы и угрозы.*

[1] Работа выполнена при финансовой поддержке Совета по грантам Президента Российской Федерации (грант МД-2209.2020.6) «Развитие системы правовых средств обеспечения кибербезопасности в Российской Федерации».

Minbaleev A. V.

THE USE OF ARTIFICIAL INTELLIGENCE AND LEGAL SUPPORT OF INFORMATION SECURITY AND CYBERSECURITY IN RUSSIA AND ABROAD: MAIN PROBLEMS

The article analyzes the main problems of ensuring information security in the Russian Federation and in the world in the process of creating and using artificial intelligence, which are a strategic direction in the field of ensuring the security of individuals, society and the state. The problem of replacing a person with artificial intelligence in the process of their own life activity is becoming very urgent today. It rises today as people continuing to make a significant number of operations for searching, processing, transmission of information becomes completely dependent on robots, trusts them more and is willing to sacrifice some of its freedoms, legitimate interests, including in the field of information security during transmission of certain kinds of human processes robots.

The article analyzes the problems of ensuring information security when using artificial intelligence in carrying out terrorist activities, as well as the use of artificial intelligence in countering terrorism. The article analyzes threats to information security in the implementation of mass surveillance and data analysis using artificial intelligence technologies, including in the process of countering crime.

Keywords: *information security, artificial intelligence, crime prevention, mass surveillance, challenges and threats.*

Современная система развития искусственного интеллекта в мире свидетельствует о стремительных темпах технологической части развития данных технологий, ее постоянном экспоненциальном росте, но при этом регулирование использование технологий искусственного интеллекта, обеспечение информационной безопасности личности, общества и государства практически находится на стадии обсуждения, а не реализации системных решений. В результате весьма актуальной становится проблема замещения искусственным интеллектом человека в процессе его собственной жизнедеятельности. Она поднимается сегодня, поскольку человек прекращая совершать значительное количество операций по поиску, обработке, передаче информации становится полностью зависимым от роботов, доверяет им все больше и готов пожертвовать рядом своих свобод, законных интересов, в том числе и в сфере информационной безопасности, в ходе передачи тех или иных видов человеческих процессов роботам. Современный мир уже столкнулся с появлением возможностей искусственного интеллекта, превышающих интеллектуальные возможности человека. В связи с этим возникает проблема обеспечения безопасности человечества (причем как информационной, так и физической) и необходимости как на техническом, правовом, так и на этическом уровне, на уровне других регуляторов предусмотреть ограничения по использованию и развитию искусственного интеллекта.

Обеспечение информационной безопасности предполагает в первую очередь введе-

ние требований об информировании тех или иных субъектов об использовании технологий искусственного интеллекта и обязательного получения письменного согласия на обработку данных с использованием таких технологий. В этой связи важно ввести в законодательстве требования о таком информировании и специальном порядке получения согласия субъектами персональных данных, а также об ответственности за не информирование, отсутствие получения согласия на обработку данных в процессе использования искусственного интеллекта [1, с. 201-204].

Сегодня активно поднимается проблема применения роботов в военных целях. Военные роботы являются одной из наиболее опасных их разновидностей и их использование должно регулироваться на международном уровне. Вопрос о регулировании использования военных роботов был поднят ещё в 2013 г. в докладе Специального докладчика ООН Кристофа Хейнса, в котором рекомендовалось: ввести национальный мораторий в отношении военных роботов; заявить в одностороннем порядке и в рамках многосторонних форумов о приверженности соблюдению норм международного гуманитарного права во всей деятельности, связанной с роботизированными системами оружия; применять строгие процедуры соблюдения данных норм на всех стадиях разработки таких систем; взять обязательство обеспечивать максимально возможную степень транспарентности применительно к своим внутренним процедурам обзора вооружений, включая параметры, используемые при испытаниях роботизированных систем [2].

Сегодня «основное направление регулирования заключается в попытках приравнять военных роботов к негуманному оружию. То есть подчинить их специальной «Конвенции о запрещении или ограничении применения конкретных видов обычного оружия, которые могут считаться наносящими чрезмерные повреждения или имеющими неизбирательное действие» 1980 г. Она ограничивает либо запрещает использование специальных видов вооружения. В частности, зажигательного оружия против населения, противопехотных мин, лазерного оружия и т.д.» [3, с. 247]. Использование роботов в качестве оружия проявляется активно в их использовании при совершении кибератак, использования в качестве информационного оружия, что требует особого международного контроля и запретов.

Противодействие терроризму. Террористы, извлекают выгоду из машинного обучения и других форм искусственного интеллекта, например, при подготовке своих военных операций и сборе информации. В частности, при проведении кибератак, автоматизированные задачи, выполняемые с использованием искусственного интеллекта, могут потенциально увеличить масштаб и влияние этих атак. ИГИЛ использовало небольшие беспилотники, вооруженные гранатами, Сирии. Тем более страшны идеи об использовании террористами «роя беспилотников» [4, 5].

Для противодействия террористам и возможным кибератакам применяется ряд мер – начиная от распознавания террористов с помощью искусственного интеллекта тех или иных лиц по камерам наблюдения, по голосу; использование искусственного интеллекта для распознавания и противодействия кибератакам со стороны террористов [6].

Использование прогнозирующих технологий искусственного интеллекта в борьбе с терроризмом часто считается пагубным воздействием на права человека, порождая спектры «докриминальных» наказаний в отношении тех или иных лиц, подозреваемых в подготовке к террористическим актам. Однако хорошо регулируемое использование новых возможностей может способствовать расширению возможностей государств по защите права граждан на жизнь при одновременном улучшении соблюдения принципов, направленных на защиту других прав человека, таких, как транспарентность, соразмерность и свобода от несправедливой дискриминации. Большинство государств сосредоточено на предотвращении террористических нападений, а не на реагировании на них. Таким образом, прогнозирование уже является центральным элементом эффективного противодействия терроризму. Искусственный интеллект позволяет анализировать большие объемы данных и может воспринимать закономерности в тех данных, которые по причинам как объема, так и размерности были бы в противном случае недоступны человеческой интерпретации. Следствием этого является то, что традиционные методы расследования, которые работают вне рамок известных подозреваемых, могут быть дополнены методами, которые анализируют деятельность широкой части всего населения для выявления ранее неизвестных угроз [7].

Противодействие терроризму, кибербе-

зопасности ставится как один из ключевых факторов, обосновывающих использование систем массового наблюдения и анализа информации о гражданах с использованием технологий искусственного интеллекта. Так, по данным Фонда Карнеги «За международный мир», как минимум 75 из 176 обследованных стран мира активно используют технологии искусственного интеллекта для целей наблюдения. К ним относятся системы распознавания лиц, интеллектуальные полицейские инструменты и создание безопасных городских платформ. Ведущими поставщиками этих систем во всем мире являются китайские фирмы, возглавляемые компанией Huawei, которая поставила эти технологии по меньшей мере в 50 государств мира [8]. Китайские компании быстро проникают на африканские рынки, предлагая правительствам льготные кредиты на покупку их оборудования и обещая создать и управлять этими системами. В Кении, например, Huawei помогла установить видеосистемы, которые развернули 1800 HD-камер и 200 HD-систем наблюдения за дорожным движением по всему Найроби [9]. В Зимбабве базирующийся в Гуанчжоу разработчик CloudWalk объявил о сделке в 2018 г. [10-12] по надзору за крупномасштабной программой распознавания лиц в сотрудничестве с властями [13]. Многие видят серьезную опасность того, что под видом борьбы с преступностью массовое наблюдение и отслеживание за гражданами, использование технологий искусственного интеллекта может подавить деятельность политической оппозиции [14].

Дебаты по поводу технологий искусственного интеллекта также происходят, когда африканские правительства и активисты сталкиваются по таким вопросам, как цифровая конфиденциальность, цензура информации, наблюдение и отключение интернета. С дефицитом законов о неприкосновенности частной жизни в таких странах, как Кения, есть озабоченность по поводу того, как правительства будут использовать эти хранилища данных, где они будут храниться, и кто будет иметь к ним доступ. Информационные и коммуникационные технологии могут быть использованы для запугивания и принуждения критиков государства. Американский аналитический центр Freedom House заявил, что Пекин обучает африканские государства некоторым из своих собственных ограничительных онлайн-мер [15].

По мере расширения масштабов деятельности Huawei в Африке в последние годы все более пристальное внимание уделяется ее деятельности. В 2018 году компания опровергла утверждения о том, что техническая инфраструктура, установленная ею в Африканском Союзе, использовалась Китаем для слежки за континентальным телом. Недавнее расследование Wall Street Journal также показало, что техники Huawei якобы помогали силам кибербезопасности в Уганде и Замбии перехватывать сообщения и выслеживать противников [16]. Стесненная в средствах полиция Уганды также купила телекамеры закрытого типа за 126 миллионов долларов у Huawei—шаг, который оппозиционные деятели опасаются использовать для идентификации и целеуказания демонстрантам и оппозиционным деятелям в преддверии выборов 2021 года [17].

Технологии искусственного интеллекта и противодействие преступности. Использование технологий искусственного интеллекта государством для реализации правоохранительной функции, в том числе при выявлении преступников и их розыске, может способствовать дискриминации отдельных социальных групп или граждан. Так, полиция США охотно использует технологии искусственного интеллекта, предназначенные для прогнозирования преступлений, чтобы решить, куда направлять офицеров для патрулирования [18]. Например, такой опыт был в Чикаго [19], где чикагская полиция использовала данные и компьютерный анализ, чтобы определить районы, в которых возможны насильственные преступления, и назначить дополнительные полицейские патрули в этих районах. Кроме того, программное обеспечение идентифицировало отдельных людей, которые, как ожидается, станут, но еще не стали жертвами или исполнителями насильственных преступлений [20]. Сегодня имеются и исследования, утверждающие, что такое расширенное использование данных о гражданах полицейскими может привести к дальнейшей ориентации на отдельные сообщества или цветных людей и подвергать их дискриминации. Анализ этих систем показывает, что данные, на которых обучаются эти системы, часто оказываются необъективными, что приводит к несправедливым результатам, таким как ложное определение того, что представители афроамериканской культуры более склонны совершать преступления, чем дру-

гие группы [18]. В последние годы имеется серьезная критика данных процессов со стороны специалистов, более 100 организаций в области гражданских прав, цифровой юстиции и общественных организаций выражают обеспокоенность по поводу досудебной оценки рисков [21].

Сегодня специалисты также приводят и ряд других случаев применения искусственного интеллекта для предупреждения и выявления преступлений [22]. Но и здесь существуют проблемы нарушений прав в сфере информационной безопасности. Таким образом, риск использования технологии искусственного интеллекта правительствами возможностями для мониторинга, отслеживания и наблюдения за отдельными людьми или социальными группами в целях ограничения или нарушения их прав реально подтверждается опытом Китая, который использует высокотехнологичные технологии искусственного интеллекта в Синьцзяне для ограничений уйгурского населения и других национальных этнических групп. Кроме того, Китай активно распространяет данные технологии, прежде всего в Африке. Правительства с демократическим режимом также могут иметь искушение злоупотреблять новыми технологическими возможностями искусственного интеллекта. Так, в США активно проводятся исследования компанией Microsoft с китайскими военными учеными об использовании искусственного интеллекта для наблюдения и анализа пространственных данных; США активно использовали технологии искусственного интеллекта для сбора и обработки данных в Европе и по всему миру в процессе разведывательной деятельности за правительствами и компаниями других стран. Это обусловливает необходимость разработки международных норм и требований к использованию технологий искусственного интеллекта в процессе наблюдения за гражданами, а также требований по использованию полученных данных для обеспечения недопустимости нарушения прав и свобод человека и гражданина, обеспечения верховенства права.

Использование технологий искусственного интеллекта государством для реализации правоохранительной функции, в том числе при выявлении преступников и их розыске, бесспорно, имеет весьма положительный опыт в США, Китае и активно распространяется в других странах. Искусственный интеллект позволяет как выявлять нарушителей, прогнозировать совершение преступлений, осуществлять розыск подозреваемых. Однако, как показывают исследования, данная практика может способствовать дискриминации отдельных социальных групп или граждан. В связи с этим необходима разработка требований о недопустимости при использовании таких технологий дискриминации отдельных социальных групп или граждан, нарушения иных прав и свобод человека и гражданина, а также требование минимизировать риск ошибки отнесения тех или иных граждан к подозреваемым в совершении преступления или к лицам, которые могут быть ошибочно отнесены к потенциальным правонарушителям. Кроме того, исследования технологий искусственного интеллекта, связанных с гендерным фактором свидетельствует, что темнокожие женщины являются наиболее неправильно классифицированной группой при распознавании, (с частотой ошибок до 34,7%). Максимальная же частота ошибок для светлокожих мужчин составляет 0,8%. Существенные различия в точности классификации более темных женщин, более светлых женщин, более темных мужчин и более светлых мужчин в системах гендерной классификации требуют безотлагательного внимания, если коммерческие компании хотят построить подлинно справедливые, прозрачные и подотчетные алгоритмы анализа лица [23].

Литература

1. Правовое регулирование цифровой экономики в современных условиях развития высокотехнологичного бизнеса в национальном и глобальном контексте : монография / под общ. ред. В. Н. Синюкова, М. А. Егоровой. Московский государственной юридический университет имени О. Е. Кутафина (МГЮА). – М.: Проспект, 2019. – 240 с.

2. Доклад Специального докладчика по вопросу о внесудебных казнях, казнях без надлежащего судебного разбирательства или произвольных казнях Кристофа Хейнса. [Электронный ресурс] – URL.: http://undocs.org/ru/A/HRC/23/47 (дата доступа 18.03.2020).

3. Аналитический обзор мирового рынка роботизации. М.: Сбербанк, 2018. С. 247

4. Renske van der Veer Terrorism in the age of technology. [Электронный ресурс] – URL.: https://www.clingendael.org/pub/2019/strategic-monitor-2019-2020/terrorism-in-the-age-of-technology/(дата доступа 18.03.2020).

5. Miles Brundage et al. The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation, (February 2018). [Электронный ресурс] – URL.: https://www.experian.co.uk/blogs/latest-thinking/data-and-innovation/ai-counter-fraud-and-the-government-opportunities-in-emerging-technologies/(дата доступа 18.03.2020).

6. Yasmin Tadjdeh Algorithmic Warfare: DoD Seeks AI Alliance to Counter China, Russia. [Электронный ресурс] – URL.: https://www.nationaldefensemagazine.org/articles/2020/3/3/algorithmic-warfare-dod-seeks-ai-alliance-to-counter-china-russia(дата доступа 18.03.2020).

7. Kathleen McKendrick Artificial Intelligence Prediction and Counterterrorism. [Электронный ресурс] – URL.: https://www.chathamhouse.org/publication/artificial-intelligence-prediction-and-counterterrorism (дата доступа 18.03.2020).

8. Steven Feldstein The Global Expansion of AI Surveillance / Carnegie Endowment for International Peace. [Электронный ресурс] – URL.: https://carnegieendowment.org/files/WP-Feldstein-AISurveillance_final.pdf (дата доступа 18.03.2020).

9. Video Surveillance as the Foundation of "Safe City" in Kenya. [Электронный ресурс] – URL.: https://www.huawei.com/en/industry-insights/technology/digital-transformation/video/video-surveillance-as-the-foundation-of-safe-city-in-kenya (дата доступа 18.03.2020).

10. Lynsey Chutel China is exporting facial recognition software to Africa, expanding its vast database. [Электронный ресурс] – URL.: https://qz.com/africa/1287675/china-is-exporting-facial-recognition-to-africa-ensuring-ai-dominance-through-diversity/ (дата доступа 18.03.2020).

11. Shan Jie China exports facial ID technology to Zimbabwe. [Электронный ресурс] – URL.: http://www.globaltimes.cn/content/1097747.shtml (дата доступа 18.03.2020).

12. Zhang Hongpei Chinese facial ID tech to land in Africa. [Электронный ресурс] – URL.: http://www.globaltimes.cn/content/1102797.shtml (дата доступа 18.03.2020).

13. Abdi Latif Dahir Chinese firms are driving the rise of AI surveillance across Africa . [Электронный ресурс] – URL.: https://qz.com/africa/1711109/chinas-huawei-is-driving-ai-surveillance-tools-in-africa/ (дата доступа 18.03.2020).

14. China's AI package for Africa includes mass surveillance technology. [Электронный ресурс] – URL.: https://mindmatters.ai/2019/05/chinas-ai-package-for-africa-includes-mass-surveillance-technology/ (дата доступа 18.03.2020).

15. Abdi Latif Dahir China is exporting its digital surveillance methods to African governments. [Электронный ресурс] – URL.: https://qz.com/africa/1447015/china-is-helping-african-countries-control-the-internet/ (дата доступа 18.03.2020).

16. Joe Parkinson, Nicholas Bariyo and Josh Chin Huawei Technicians Helped African Governments Spy on Political Opponents. [Электронный ресурс] – URL.: https://www.wsj.com/articles/huawei-technicians-helped-african-governments-spy-on-political-opponents-11565793017 (дата доступа 18.03.2020).

17. Musa Okwonga On returning to Uganda, Museveni's staying power and the significance of Bobi Wine. [Электронный ресурс] – URL.: https://qz.com/africa/1631116/returning-to-uganda-musevenis-reign-and-bobi-wines-music/ (дата доступа 18.03.2020).

18. Why big-data analysis of police activity is inherently biased. [Электронный ресурс] – URL.: https://theconversation.com/why-big-data-analysis-of-police-activity-is-inherently-biased-72640&usg=ALkJrhhJAILrqlCIewb2Kby9hP_4t5a5Dg (дата доступа 18.03.2020).

19. Электронный ресурс. – URL.: https://www.nbcnews.com/news/us-news/chicago-police-department-goes-high-tech-fight-rise-killings-n713206&usg=ALkJrhhHo866Kk6_sm86Gy6WOtFAtZiq_A (дата доступа 18.03.2020).

20. Andrew V. Papachristos CPD's crucial choice: Treat its list as offenders or as potential victims? [Электронный ресурс] – URL.: https://www.chicagotribune.com/opinion/commentary/ct-gun-violence-list-chicago-police-murder-perspec-0801-jm-20160729-story.html&usg=ALkJrhhGNFNEg9CVZfpdgbscS95C0X4rZA (дата доступа 18.03.2020).

21. How well do IBM, Microsoft, and Face++ AI services guess the gender of a face?[Электронный ресурс] – URL.: https://z5h64q92x9.net/proxy_u/en-ru.ru/gendershades.org/ (дата доступа 18.03.2020).

22. More than 100 Civil Rights, Digital Justice, and Community-Based Organizations Raise Concerns About Pretrial Risk Assessment. [Электронный ресурс] – URL.: https/civilrights.org/2018/07/30/more-than-100-civil-rights-digital-justice-and-community-based-organizations-raise-concerns-about-pretrial-risk-assessment/ ; Karen Hao AI is sending people to jail—and getting it wrong, Jan 21, 2019 // https://www.technologyreview.com/s/612775/algorithms-criminal-justice-ai/ (дата доступа 18.03.2020).

23. Daniel Faggella AI for Crime Prevention and Detection – 5 Current Applications. [Электронный ресурс] – URL.: https://emerj.com/ai-sector-overviews/ai-crime-prevention-5-current-applications/ (дата доступа 18.03.2020).

References

1. Pravovoye regulirovaniye tsifrovoy ekonomiki v sovremennykh usloviyakh razvitiya vysokotekhnologichnogo biznesa v natsional'nom i global'nom kontekste : monografiya / pod obshch. red. V. N. Sinyukova, M. A. Yegorovoy. Moskovskiy gosudarstvennoy yuridicheskiy universitet imeni O. Ye. Kutafina (MGYUA). – M.: Prospekt, 2019. – 240 s.

2. Doklad Spetsial'nogo dokladchika po voprosu o vnesudebnykh kaznyakh, kaznyakh bez nadlezhashchego sudebnogo razbiratel'stva ili proizvol'nykh kaznyakh Kristofa Kheynsa. [Elektronnyy resurs] – URL.: http://undocs.org/ru/A/HRC/23/47 (data dostupa 18.03.2020).

3. Analiticheskiy obzor mirovogo rynka robotizatsii. M.: Sberbank, 2018. S. 247.

МИНБАЛЕЕВ Алексей Владимирович, заведующий кафедрой информационного права и цифровых технологий Московского государственного юридического университета имени О.Е. Кутафина (МГЮА), профессор кафедры теории государства и права, конституционного и административного права Южно-Уральского государственного университета (национального исследовательского университета), доктор юридических наук, доцент. 123001, г. Москва, ул., Садовая-Кудринская, 9. 454080, г. Челябинск, пр. Ленина, 76. Email: alexmin@bk.ru

MINBALEEV Aleksey, head. Department of information law and digital technologies of the Moscow state law University named after O. E. Kutafin (MSAL), Professor of the Department of theory of state and law, constitutional and administrative law, South Ural state University (national research university) Doctor of Law, Associate Professor. 123001, г. Москва, ул., Садовая-Кудринская, 9. 454080, г. Челябинск, пр. Ленина, 76. Email: alexmin@bk.ru

УДК 004.056.53

Вестник УрФО № 1(35) / 2020, с. 72–76

Мухачев С. В.

DOI: 10.14529/secur200108

ИНЦИДЕНТЫ, СВЯЗАННЫЕ С ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТЬЮ, НА ОБЪЕКТАХ ЯДЕРНОЙ ИНФРАСТРУКТУРЫ

В статье рассмотрена хронология и особенности инцидентов, связанных с информационной безопасностью, на объектах ядерной инфраструктуры. Приводится краткое описание инцидентов в различных странах, мотивы атак, предпринятые для обезвреживания меры и последствия. Выделены особенности, присущие каждой из приведенных кибератак. Показано, что проблема существования инцидентов, связанных с информационной безопасностью, на объектах ядерной инфраструктуры возникла с развитием и внедрением информационных и управляющих систем с использованием компьютерной техники. Это ясно просматривается при анализе хронологии описываемых инцидентов. Принятие федерального закона Российской Федерации «О безопасности критической информационной инфраструктуры Российской Федерации» позволяет правовыми средствами вести борьбу с кибератаками на объектах ядерной инфраструктуры.

Ключевые слова: *информационная безопасность, критическая информационная инфраструктура, ядерный объект, кибератака, вирус.*

Mukhachev S.V.

INCIDENTS RELATED TO INFORMATION SECURITY AT NUCLEAR INFRASTRUCTURE OBJECTS

The article discusses the chronology and features of incidents related to information security at nuclear infrastructure facilities. A brief description of incidents in various countries, the motives of the attacks, measures taken to neutralize the measures and consequences are given. The features inherent in each of the cyberattacks are highlighted. It is shown that the problem of the existence of incidents related to information security at nuclear infrastructure facilities arose with the development and implementation of information and control systems using

computer technology. This is clearly seen in the analysis of the chronology of the described incidents. The adoption of the federal law of the Russian Federation "On the security of critical information infrastructure of the Russian Federation" allows legal means to combat cyber attacks on nuclear infrastructure facilities.

Keywords: *information security, critical information infrastructure, nuclear facility, cyber attack, virus.*

Компьютерные атаки могут нанести вред не только физически лицам владельцам компьютеров, как это было на заре развития компьютерной и информационно-телекоммуникационной техники, но и объектам критической информационной инфраструктуры информационным системам, информационно-телекоммуникационным сетям, автоматизированным системам управления субъектов критической информационной инфраструктуры. Это могут быть промышленные технологические системы, системы жизнеобеспечения городов, системы управления объектами ядерной инфраструктуры и другие объекты, относящиеся к критической информационной инфраструктуре. Сбои в работе таких систем могут нести пагубные или даже катастрофические последствия.

Современный мир сосредоточен на проблеме обеспечения безопасности инфраструктуры и информационных систем. Частные компании и государственные структуры прогнозируют потенциальные убытки, которые могут быть причинены в случае компьютерной атаки. Причем, часто речь идет об объектах инфраструктуры, от которых напрямую зависит жизнедеятельность целых городов, отдельных регионов и стран. Чтобы противостоять существующим и вновь возникающим угрозам, разрабатываются методы борьбы с ними.

В ответ на существующие угрозы, с целью защиты объектов такой инфраструктуры, в Российской Федерации принят Федеральный закон от 26.07.2017 № 187-ФЗ «О безопасности критической информационной инфраструктуры Российской Федерации» [1].

В перечень объектов критической информационной инфраструктуры включены информационные системы, информационно-телекоммуникационные сети, автоматизированные системы управления, функционирующие в сфереатомной энергетики.

История вопроса и особенности, присущие объектам ядерной инфраструктуры, освещаются в ряде публикаций, посвященных данной тематике. С точки зрения рассматриваемой темы интересен доклад Королевского института международных отношений ChathamHouse «Кибербезопасность на объектах гражданской атомной инфраструктуры: понимание рисков» [2]. В докладе, опубликованном в 2015 году, подвергаются оценке возможные уязвимости и методы укрепления кибербезопасности. Эксперты провели состояния защищенности ядерных объектов, привели сведения о 50-ти киберинцидентах в различных странах мира.

Обратившись к истории вопроса, можно найти ряд примеров реализации киберугроз на объектах ядерной инфраструктуры. Остановимся на них более подробно, выстроив хронологию, и обратим внимание на особенности инцидентов. Следует обратить внимание на то, что далеко не все инциденты такого рода становятся известны общественности. Компании не заинтересованы в их разглашении по многим причинам: раскрытие особенностей системы безопасности, нежелание имиджевых потерь, разглашение технических секретов и т.д.

Один из первых известных инцидентов произошел в 1995 году на Игналинской атомной электростанции (г. Висагинас, Литва). Произошло заражение вирусом программного обеспечения для управления процессом перезагрузки ядерного топлива. Общепринятая версия связывает инцидент с местной преступной группировкой, которая таким образом пыталась отомстить за жесткий судебный приговор одному из участников группировки[3]. Вредоносная программа была внесена сообщником, который являлся сотрудником электростанции и был занят обслуживанием системы управления станции. Этот инцидент, имеющий связь с криминальным миром, можно отнести к кибертерроризму. Факт заражения был вовремя обнаружен, вирус локализован. Каких-либо последствий заражение не имело.

Другойинцидентимел место в 1998 году, когда кибератаке подвергся индийский Центр ядерных исследований им. Хоми Баба (BabhaAtomicResearchCenter) (Индия). Там террористы угрожали вывести из строя систему управления реактором. Насколько ре-

альнадля исполнения была угроза, в докладе не сообщается. Однако следует обратить внимание на то обстоятельство, что даже в случае угрозы без реальной возможности реализовать кибератаку, не исключена внеплановая остановка ядерно-опасных работ. Кроме того, необходимые проверки всех систем занимают очень продолжительное время (до нескольких месяцев). Поэтому даже имитация кибератаки (например, заражения вирусом систем объекта ядерной инфраструктуры) ведет к серьезным потерям времени, финансовым издержкам, нарушениям сроков выполнения работ.

Примечателен инцидент, зафиксированный в США в 2003 году. На атомной электростанции «Дэвис-Бессе» (Davis-Besse) в американском штате Огайо инженер получил доступ к оборудованию станции с домашнего компьютера, зараженного вирусом Slammer (сетевой червь, вызывающий отказ в обслуживании хостов в Интернете и сильное снижение общего интернет-трафика). Две системы управления оказались поражены вирусом, одна из которых отвечала за мониторинг безопасности. Она была отключена в течение пяти часов. Нужно признать, что в этом случае реальной опасности серьезных последствий, связанных с ядерными технологиями, не было, так как реактор в Дэвиса-Бессе не работал: он был закрыт почти два года после обнаружения отверстия в корпусе реактора.

Еще одна скандальная кибератака на ядерный объект произошла в 2010 году. Вирус Stuxnet, проникший в систему управления центрифугами на заводе по обогащению ядерного топлива в Иране (г. Натанз), парализовал деятельность завода [2,4]. В тот момент на заводе работало 18 каскадов по 164 центрифуги в каждом. В них было загружено около 1240 кг гексафторида урана [5]. Stuxnetдействовал таким образом, что оборудование переходило в нештатный аварийный режим. В результате кибератаки значительная часть центрифуг разрушились. Большинство исследователей считает, что оригинальный компьютерный вирус, получивший название Stuxnet был разработан спецслужбами США и Израиля для уничтожения иранской ядерной программы (официального подтверждения эта версия не имеет, что, впрочем, понятно). Он был создан специально для работы с компьютерами строго определенной конфигурации, то есть имел узкую специализацию. Попав в компьютер, Stuxnet сканировал установленное на нем программное обеспечение, чтобы определить, входит ли данный компьютер в автоматизированную систему управления центрифугами. Такие системы строго специфичны для каждого завода и имеют уникальные системы: датчиков, управления узлами и агрегатами. Stuxnet искал строго определённую цель. Если компьютер был нецелевым, то вирус не проявлял себя и ожидал возможности переместиться далее.

Таким образом, данный инцидент связан с противоборством нескольких государств, а в качестве оружия был применен «боевой» вирус Stuxnet.

В 2016 году обнаружен вирус в системе управления реактором на атомной электростанции Kernkraftwerk Gundremmingen в Германии[6,7]. В ходе плановой проверки в компьютере системы управления тепловыделяющей сборки был обнаружен вирус. Согласно пресс-релизу компании, он был выявлен при проверке съемных носителей данных и устройств программного управления. По информации энергокомпании RWE, которая эксплуатирует ядерный объект, инцидент не создал угрозу безопасности персоналу предприятия и местному населению. Однако надзорные ведомства и Федеральное управление по информационной безопасностибыли проинформированы о случившемся.

В июне 2017 года вирус-вымогатель Petya. А поразил системы радиационного мониторинга и электронный документооборотЧернобыльской атомной электростанции[8].Данная вредоносная программасетевой червь с функциями программы-вымогателя.Она поражает компьютеры под управлением операционной системы Microsoft Windows. Первые разновидности вируса были обнаружены в марте 2016 года. Программа шифруетфайлы нажёстком дискекомпьютера-жертвы, а также перезаписывает и шифрует главную загрузочную запись данные, необходимые для загрузки операционной системы. После срабатывания Petya. Афайлы, хранящиеся на компьютере, становятся недоступными. После этого программа-вымогатель требует выкуп –эквивалент трехсот долларов в биткойнах за расшифровку и восстановление доступа к файлам. Причем, как заявили исследователи «Лаборатории Касперского» после того, как файлы зашифрованы, уже нет возможности расшифровать их [9]. Усилиями персонала удалось справиться с вирусом.

Проблема существования инцидентов, связанных с информационной безопасностью, на объектах ядерной инфраструктуры возникла с развитием и внедрением информационных и управляющих систем с использованием компьютерной техники. Это ясно просматривается при анализе хронологии описываемых инцидентов.

Из описания приведенных инцидентов можно видеть, что угрозы кибер-атак на объекты ядерной инфраструктуры достаточно реальны и относительно многочисленны.

Разнообразны мотивы кибератак на объекты ядерной критической инфраструктуры: борьба государств между собой, терроризм, корыстные побуждения, халатность.

Различны и последствия кибератак. Понятно, что они могут быть очень серьезными, вплоть до катастрофических. В случае серьезного инцидента возможны ядерные аварии с выходом больших доз излучения и загрязнение местности и атмосферы радиоактивными элементами.

Обладание ядерными технологиями является сегодня серьезным политическим аргументом. Поэтому ведется борьба между государствами за право обладания ими. И как последствие такой борьбы уничтожение оборудования для производства ядерных материалов и их использования. Как раз такой случай и имел место в Иране.

В других обсуждаемых инцидентах обошлось без серьезных последствий. Конечно же, защите объектов ядерной критической инфраструктуры всегда уделялось серьезное внимание. Но, тем не менее, как следует из вышеизложенного, инциденты, связанные с кибератаками, до сих пор имеют место. Обусловлено это, прежде всего, стремительным развитием информационных технологий, изменением управляющих программно-аппаратных комплексов. Вместе с этим возникают недокументированные возможности и ошибки, что и обусловливает появлением новых возможностей для организации атак.

Объекты ядерной критической инфраструктуры требуют повышенного внимания ккиберугрозам. Поэтому меры, предпринимаемые государством в сфере защиты объектов критической информационной инфраструктуры крайне важны и актуальны и для защиты объектов ядерной инфраструктуры. Федеральный закон № 187-ФЗ «О безопасности критической информационной инфраструктуры Российской Федерации» предъявляет ряд серьезных требований ксубъектам критической информационной инфраструктуры с целью предотвращения киберугроз: определение перечня объектов; предоставление утвержденного перечня объектов во ФСТЭК России; категорирование объектов; предоставление во ФСТЭК России сведений о результатах категорирования объектов. Этот же нормативный актобусловил создание Государственной системы обнаружения, предупреждения и ликвидации последствий компьютерных атак (ГосСОПКА), которая предназначена для сбора и обмена информацией о компьютерных атаках на территории РФ. Предпринятые меря должны стать серьезным шагом к предотвращению киберугроз на объектах ядерной инфраструктуры.

Литература

1. Федеральный закон № 187-ФЗ от 26.07.2017 «О безопасности критической информационной инфраструктуры Российской Федерации» [Электронный ресурс]. Режим доступа: URL:http://www.consultant.ru (дата обращения 05.04.2020).

2. Cyber Security at Civil Nuclear Facilities Understanding the Risks. Chatham House Report Caroline Baylon with Roger Brunt and David Livingstone September, 2015.

3. Вакцина для атома: кибербезопасность АЭС [Электронный ресурс]. Режим доступа: URL: https://russiancouncil.ru/analytics-and-comments/analytics/vaktsina-dlya-atoma-kiberbezopasnost-aes/#7(дата обращения 05.04.2020).

4. Вирус страшнее бомбы. Как хакеры уничтожили ядерный завод в Иране [Электронный ресурс]. Режим доступа: URL: https://life.ru/p/1047800(дата обращения 05.04.2020).

5. Ядерное нераспространение: краткая энциклопедия / [И. А. Ахтамзян и др.]; гл. ред. А. В. Хлопков. Москва: РОССПЭН, 2009.

6. Киберугрозы: защищены ли АЭС? [Электронный ресурс]. Режим доступа: URL: http://atomicexpert.com/page1081638.html (дата обращения 05.04.2020).

7. В компьютере на баварской АЭС обнаружен вирус[Электронный ресурс]. Режим доступа: URL:https://www.dw.com/ru/в-компьютере-на-баварской-аэс-обнаружен-вирус/a-19216272-0 (дата обращения 05.04.2020).

8. Вирус-вымогатель Petya поразил Чернобыльскую АЭС [Электронный ресурс]. Режим доступа: URL: https://ria.ru/20170627/1497397238.html (дата обращения 05.04.2020).

9. Петя стирает память[Электронный ресурс]. Режим доступа: URL: https://www.gazeta.ru/tech/2017/06/29/10752467/petya_the_destroyer.shtml (дата обращения 05.04.2020).

References

1. Federal'nyy zakon № 187-FZ ot 26.07.2017 «O bezopasnosti kriticheskoy informatsionnoy infrastruktury Rossiyskoy Federatsii» [Elektronnyy resurs]. Rezhim dostupa: URL:http://www.consultant.ru (data obrashcheniya 05.04.2020).

2. Cyber Security at Civil Nuclear Facilities Understanding the Risks. Cha-tham House Report Caroline Baylon with Roger Brunt and David Livingstone September, 2015.

3. Vaktsina dlya atoma: kiberbezopasnost' AES [Elektronnyy resurs]. Rezhim dostupa: URL: https://russiancouncil.ru/analytics-and-comments/analytics/vaktsina-dlya-atoma-kiberbezopasnost-aes/#7(data obrashcheniya 05.04.2020).

4. Virus strashneye bomby. Kak khakery unichtozhili yadernyy zavod v Irane [Elektronnyy resurs]. Rezhim dostupa: URL: https://life.ru/p/1047800(data obrashcheniya 05.04.2020).

5. Yadernoye nerasprostraneniye: kratkaya entsiklopediya / [I. A. Akhtamzyan i dr.]; gl. red. A. V. Khlopkov. Moskva: ROSSPEN, 2009.

6. Kiberugrozy: zashchishcheny li AES? [Elektronnyy resurs]. Rezhim dostupa: URL: http://atomicexpert.com/page1081638.html (data obrashcheniya 05.04.2020).

7. V komp'yutere na bavarskoy AES obnaruzhen virus[Elektronnyy re-surs]. Rezhim dostupa: URL:https://www.dw.com/ru/v-komp'yutere-na-bavarskoy-aes-obnaruzhen-virus/a-19216272-0 (data obrashcheniya 05.04.2020).

8. Virus-vymogatel' Petya porazil Chernobyl'skuyu AES [Elektronnyy resurs]. Rezhim dostupa: URL: https://ria.ru/20170627/1497397238.html (data obrashcheniya 05.04.2020).

9. Petya stirayet pamyat'[Elektronnyy resurs]. Rezhim dostupa: URL: https://www.gazeta.ru/tech/2017/06/29/10752467/petya_the_destroyer.shtml (data obrashcheniya 05.04.2020).

МУХАЧЕВ Сергей Валентинович, кандидат физико-математических наук, доцент, доцент кафедры информационных технологий и защиты информации, Уральский государственный университет путей сообщения.620034, г. Екатеринбург, ул. Колмогорова, 66. E-mail: msv62@yandex.ru

MUKHACHEV Sergey, candidate of physical and mathematical Sciences, associate Professor, associate Professor of the Department of information technology and information security, Ural State University of Railway Transport.

УДК 004.85 Вестник УрФО № 1(35) / 2020, с. 77–83

Асяев Г. Д., Соколов А. Н.

DOI: 10.14529/secur200109

ОБНАРУЖЕНИЕ ВТОРЖЕНИЙ НА ОСНОВЕ АНАЛИЗА АНОМАЛЬНОГО ПОВЕДЕНИЯ ЛОКАЛЬНОЙ СЕТИ С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ С УЧИТЕЛЕМ

В работе представлены модели процесса обнаружения вторжений, построенные на основе трёх методов машинного обучения: метода деревьев решений, метода ближайших соседей и метода случайного леса. Основной задачей при моделировании является классификация состояний автоматизированной системы управления (АСУ) (аномальное, нормальное). Рассмотрены параметры, влияющие на обнаружение аномального поведения: протокол, сервисные данные, используемые флаги, количество неудачных попыток входа, продолжительность атаки. Для моделирования процесса поиска аномалий выбран набор данных транспортно-сетевого уровня АСУ, состоящий из необработанных дампов TCP/IP в ситуации, когда сеть подверглась множественным атакам. Для каждого соединения TCP/IP фиксировались 3 качественных и 38 количественных признаков, среди которых выделены наиболее важные признаки, влияющие на обучение. Прогнозирование ответа проводилось на контрольной (тестовой) выборке. Основными критериями выбора математической модели для решаемой задачи являлись количество правильно распознанных (accuracy) аномалий, точность (precision) и полнота (recall) ответов. На основании проведенных исследовании был выбран оптимальный алгоритм для обнаружения аномалий.

Ключевые слова: *автоматизированная система управления, обнаружение вторжений, уязвимость, аномалия, метод машинного обучения, метод деревьев решений, метод ближайших соседей, метод случайного леса.*

DETECTION OF INVASION ON THE BASIS OF ANALYSIS OF ANOMALOUS BEHAVIOR OF A LOCAL NETWORK USING MACHINE-LEARNING ALGORITHMS WITH A TEACHER

The paper presents models of the intrusion detection process based on three machine learning methods: the decision tree method, the nearest neighbor method and the random forest method. The main task in modeling is to classify the ACS states (abnormal, normal). Parameters affecting the detection of anomalous behavior are considered: protocol, service data, flags used, number of unsuccessful attempts to enter, duration of the attack. To simulate the process of anomaly detection, the data set of the transport and network level of the control system, consisting of raw TCP/IP dumps in a situation where the network has been subjected to multiple attacks, was selected. For each TCP/IP connection, 3 qualitative and 38 quantitative features were recorded, among which the most important features affecting the learning were highlighted. The response was predicted in a control (test) sample. The main criteria for choosing a mathematical model for the task were the number of correctly recognized (accuracy) anomalies, accuracy (precision) and completeness (recall) of answers. The optimal algorithm for detection of anomalies was chosen on the basis of the conducted research.

Keywords: *automated control system, intrusion detection, vulnerability, anomaly, machine learning method, decision tree method, closest neighbour method, random forest method..*

Автоматизированные системы управления (АСУ) широко используется во многих отраслях производственной деятельности. Через их элементы управления проходят большие объемы данных (big data), основной угрозой для которых является вмешательство террористических, экстремистских и враждебно настроенных групп в управление автоматизированными системами, в том числе с целью вывода их из строя [1]. Количество примеров подобных атак (Stuxnet, Crouching Yeti, BlackEnergy и т.п.) ежегодно растет вследствие большого числа уязвимостей у эксплуатируемых систем. Влияние уязвимости на вероятность осуществления атаки тем выше, чем [1]:

• большее число узлов, реализующих функцию, подвержено уязвимости;

• большее число функций обслуживается уязвимым программным обеспечением.

Для обеспечения безопасности информации АСУ необходимо обеспечение непрерывного контроля трафика всех взаимодействий системы. Постоянный анализ этих данных позволяет своевременно выявлять аномальное поведение системы, связанное с её некорректным функционированием. Наиболее критичной является задача сохранения способности АСУ к корректному функционированию в условиях деструктивных информационных воздействий. Успешная реализация кибератак на такие системы может повлечь за собой негативные финансовые последствия, экологические катастрофы или даже привести к гибели людей. Поэтому важными особенностями АСУ являются [2]:

• непрерывный режим работы: остановка работы АСУ, как правило, либо невозможна, либо влечет за собой значительные финансовые потери;

• в зависимости от особенностей технологического процесса различные элементы автоматизированной информационной системы вносят различный вклад по степени критичности возможного ущерба.

Для поддержания системы информационной безопасности АСУ на требуемом уровне, необходима регулярная установка критических обновлений, направленных на исправление уязвимостей. Тем не менее, в силу вышеуказанных особенностей АСУ, это не всегда рационально [3].

сировались 3 качественных и 38 количественных признаков. При анализе данных использованы следующие признаки (см. рис. 1):

• duration – продолжительность атаки;

• protocol_type – используемый протокол;

• service – сервисные данные;

• flag – используемые флаги;

• num_failed_logins – количество неудачных попыток входа в систему;

• logged_in – количество вхождений в систему;

• class – критерий, характеризующий поведение системы как нормальное или аномальное.

На рис. 2 приведена диаграмма, показывающая соотношение нормального и ано-

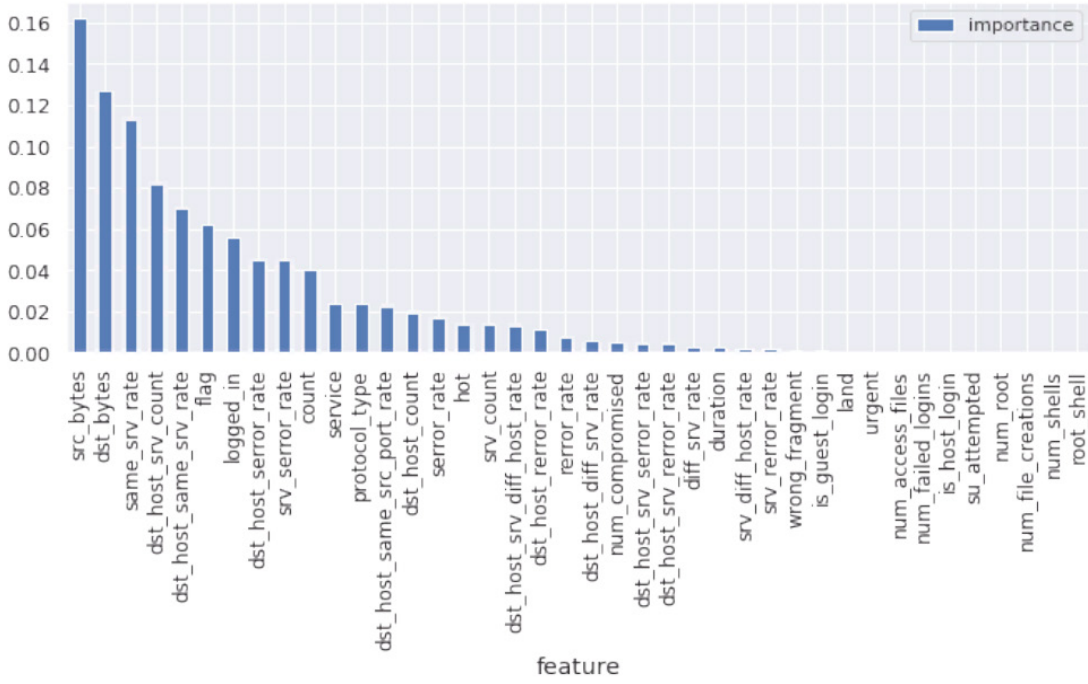
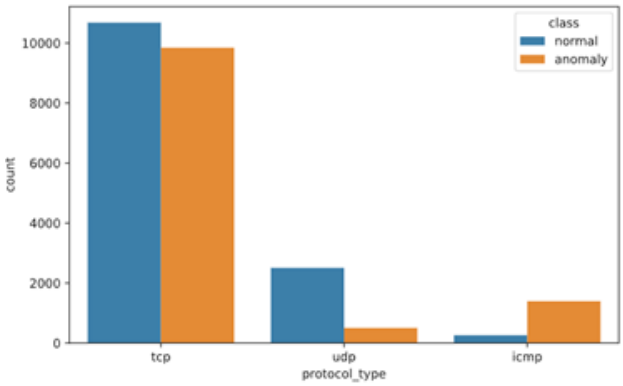


Рис. 1. Диаграмма значимости признаков по степени их влияния на обучение

Для моделирования процесса поиска аномалий выбран набор данных транспортно-сетевого уровня АСУ, состоящий из необработанных дампов TCP/IP в ситуации, когда сеть подверглась множественным атакам. Каждое соединение состоит из последовательности TCP-пакетов, начинающихся и заканчивающихся в моменты времени, в промежутке между которыми данные по определенному протоколу передаются на IP-адрес источника, а с него – на целевой IP-адрес. Кроме того, каждое соединение помечается как нормальное или как атака определенного типа. Размер каждой записи соединения – около 100 байт. Для каждого соединения TCP/IP фик-

мального состояния признака class для трёх типов протоколов (TCP, UDP и ICMP). Из рисунка видно, что наибольшее общее число аномальных состояний возникает при передаче данных с использованием протокола TCP (9845 или 48,0 %). Однако можно заметить, что злоумышленник активно использует протокол межсетевых управляющих сообщений ICMP, так как количество срабатываний счетчика аномалий примерно в 5 раз выше по сравнению с нормальным режимом работы (1394 или 84,2 %) [4]. Наименьшее количество аномалий наблюдается при использовании протокола UDP (504 или 16,7 %).

Диаграммы распределений аномального



class	anomaly	normal	All
protocol_type			
icmp	1394	261	1655
tcp	9845	10681	20526
udp	504	2507	3011
All	11743	13449	25192

Рис. 2. Соотношение нормального и аномального состояния признака class для протоколов TCP, UDP и ICMP

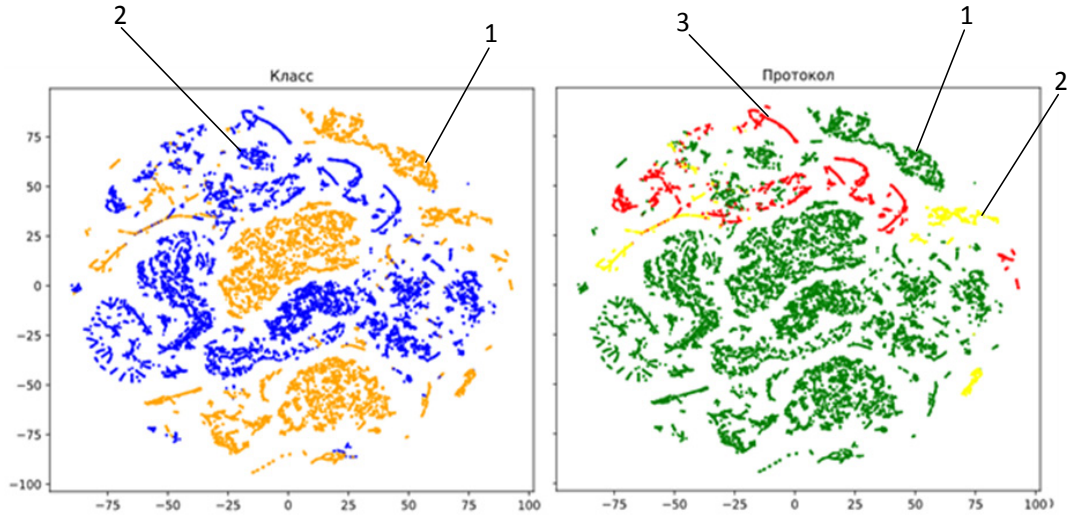


Рис. 3. Распределения аномального (оранжевый, 1) и нормального (синий, 2) состояний системы в зависимости от используемого протокола: TCP (зеленый, 1), UDP (желтый, 2) и ICMP (красный, 3)

и нормального состояний системы в зависимости от типов протоколов приведены на рис. 3.

Моделирование процесса поиска аномалий проведено с использованием трёх методов машинного обучения: метода деревьев решений, метода ближайших соседей и метода случайного леса. Основной задачей при моделировании является классификация состояния системы (аномальное, нормальное). Моделирование проводилось в среде Jupyter Notebook с использованием библиотек scikit-learn, pandas, numpy. Для обучения математических моделей категориальные признаки были закодированы с использованием метода «one-hot encoding».

Метод деревьев решений. Разделим обучающуюся выборку в соотношении 70/30, предварительно перемешав строки для оптимального обучения на соответствующей выборке. Размер обучающейся выборки равен 17634, при этом количество уникальных признаков равно 39. Важным параметром для метода деревьев решений является глубина дерева. Уменьшение глубины дерева приводит к недообучению, в то время как увеличение глубины может переобучить сеть, а математическая модель будет некорректно работать с новыми данными [5].

Значение кросс-валидации выберем равным 5, а глубину дерева 4. Точность распознанных аномалий при этом составила 97,72 %. Лучше всего эту модель удалось обучить при глубине дерева, равной 10, и параметре кросс-валидации, равному 10. Количество признаков, по которым произошло наилучшее разбиение в дереве, равно 25. Ниже представлены оптимальные параметры дерева, применяемые в обучении этой модели:

criterion = '**gini**', splitter = '**best**', max_depth = **10**, min_samples_split = **2**, min_samples_leaf = **1**, min_weight_fraction_leaf = **0.0**, max_features = **25**, random_state = **17**, max_leaf_nodes = **None**, min_impurity_decrease = **0.0**, min_impurity_split = **None**, class_weight = **None**, presort = **False**.

Выборка была разделена на обучающую и тестовую. Precision = 92,22%, recall =91,08%. Доля распознанных аномалий на обучающей выборке составила 99,56 %, что свидетельствует о хорошей степени обученности модели (см. рис. 4).

На вход обученной модели подавалась тестовая выборка. Доля правильных ответов на тестовой выборке составило 99,02 %. Таким образом, представленная математическая модель достаточно хорошо распознает аномалии.

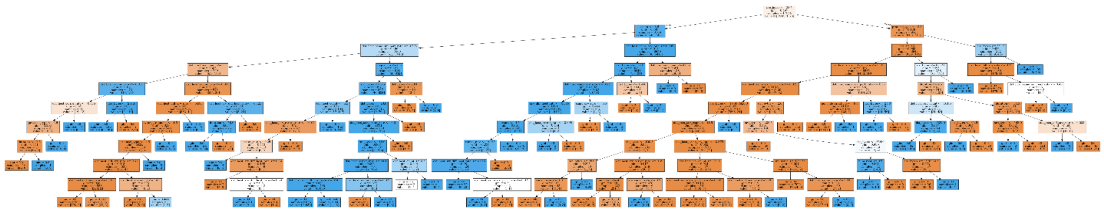


Рис. 4. Дерево решений обнаружения аномалий

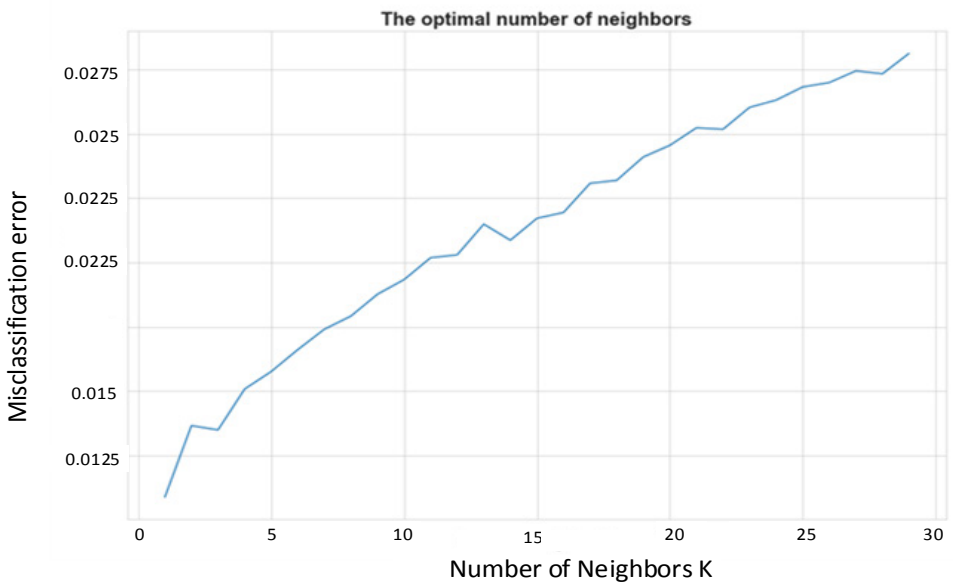


Рис. 5. Зависимость числа ближайших соседей от среднеквадратичной ошибки

Метод ближайших соседей. Данный метод относится к методам классификации без учителя [6]. Данные группируются по схожему признаку на основе рассчитанных весов. Для обучения модели выбрано значение кросс-валидации, равное 10, как и в методе деревьев решений. Максимальную долю правильных ответов удалось получить при числе ближайших соседей, равным 1. Доля правильных ответов составила 98,91 %. Ниже представлены оптимальные параметры, применяемые в обучении данной модели методом ближайших соседей (см. рис. 5):

n_neighbors = **1**, weights = '**uniform**', algorithm = '**auto**', leaf_size = **30**, p = **2**, metric = '**minkowski**', metric_params = **None**, n_jobs = **None**.

На тестовой выборке количество правильно распознанных аномалий составило 99,03 %, что соизмеримо с результатами, полученными с использованием метода деревьев решений. Precision = 93,92%, recall = 95,78%.

Метод случайного леса (Random forest). Этот метод построен на совокупности деревьев решений, прогноз которых усредняется. Такой подход имеет очевидные преимущества, связанные с повышенной точностью прогноза и минимизация процесса обучения [7]. Однако, при этом он требует значительных вычислительных мощностей. Максимальную точность прогноза удалось получить при максимальной глубине дерева, равной 10, и количестве признаков, равным 11. Точность прогноза составила 99,64 %. Precision = 98,29%, recall =98,68%.

На тестовой выборке количество правильно распознанных аномалий составило 99,69 %, что является лучшим результатом среди представленных:

RandomForestClassifier(n_estimators =

'**warn**', criterion = '**gini**', max_depth = **10**, min_samples_split = **2**, min_samples_leaf = **1**, min_weight_fraction_leaf = **0.0**, max_features = **11**, max_leaf_nodes = **None**, min_impurity_decrease = **0.0**, min_impurity_split = **None**, bootstrap = **True**, oob_score = **False**, n_jobs = **None**, random_state = **None**, verbose = **0**, warm_start = **False**, class_weight = **None**).

Специфика работы АСУ, как правило такова, что остановка её функционирования с целью обновления систем безопасности [8], обновления сигнатур антивирусных баз данных, может повлечь за собой значительные финансовые потери. Одним из решений, минимизирующим необходимость такой остановки, является анализ сетевого трафика с использованием методов машинного обучения. Анализ трафика позволяет распознавать аномалии в режиме реального времени, и, на основании этого, обнаруживать уязвимости системы. Обнаруженные уязвимости определяют необходимость установки критических обновлений. Наилучшие результаты среди рассмотренных получены с использованием метода случайного леса (Random forest): количество правильно распознанных аномалий составило 99,69 %, а полнота и точность (многообразие правильно распознанных ответов и точность отнесения к аномалии) составила 98%.

Статья выполнена при поддержке Правительства РФ (Постановление №211 от 16.03.2013 г.), соглашение № 02.A03.21.0011.

Литература

1. Андрей Заикин. Почему защита АСУ ТП сегодня стала критически важной? [Электронный ресурс] //https://www.securitylab.ru/analytics/484730.php. (Дата обращения: 10.03.2020).

2. A. Mansouri, B. Majidi and A. Shamisa, "Anomaly detection in industrial control systems using evolutionary-based optimization of neural networks", Communications on Advanced Computational Science with Applications, vol. 2017, no. 1, pp. 49–55. Available: 10.5899/2017/cacsa-00074.

3. А.Е. Баринов, С.В. Скурлаев, А.Н. Соколов. "Методика оценки рисков, вызванных уязвимостями в программном обеспечении автоматизированных систем управления технологическими процессами", Вестник УрФО. Безопасность в информационной сфере., № 3 (25), с. 34–42, 2017.

4. C. Feng, T. Li, D. Chana, "Multi-level anomaly detection in industrial control systems via package signatures and lstm networks", Dependable Systems and Networks (DSN), 47th Annual IEEE/IFIP International Conference on. – IEEE, vol. 47, pp. 261–272, 2017.

5. А.А. Бранитский, И.В. Котенко, "Анализ и классификация методов обнаружения сетевых атак", Труды СПИИРАС, № 45, с. 207–244, 2016.

6. S. N. Shirazi, "Evaluation of anomaly detection techniques for scada communication resilience", Resilience Week (RWS), IEEE, pp. 140–145, 2016.

7. М.С. Пырьев, А.С. Коллеров, "Средства анализа сетевого трафика локальной вычислительной сети в ретроспективе", Вестник УрФО. Безопасность в информационной сфере, (4(34)), с. 58–62, 2019.

8. M. Chandrashekar, Y. Lee and D. Medhi, "Real-time network anomaly detection system using machine learning", 11th International Conference on the Design of Reliable Communication Networks (DRCN), Kansas City, MO, vol. 11, pp. 267–270, 2015.

References

1. Andrey Zaikin. Why is the protection of process control systems now critical? [Electronic resource] // https://www.securitylab.ru/analytics/484730.php. (Date of treatment: 10.03.2020).

3. A. E. Barinov, S. V. Skurlaev, A. N. Sokolov, "Methodology for assessing the risks caused by vulnerabilities in the software of automated process control systems", Bulletin of the Urals Federal District. Security in the information field., vol. 3(25), pp. 34-42, 2017.

5. A.A. Branitsky, I.V. Kotenko, "Analysis and classification of network attack detection methods", Tr. SPIIRAS, vol. 45, pp. 207–244, 2016.

7. M.S. Pyryev, A.S. Kollerov, "A retrospective analysis of the network traffic of a local computer network", Proceedings of the XVIII All-Russian Scientific and Practical Conference of Students, Graduate Students and Young Scientists "Information Space Security". – Magnitogorsk, vol. 18, pp.. 302–306, 2019.

АСЯЕВ Григорий Дмитриевич, аспирант кафедры защиты информации ФГАОУ ВО «Южно-Уральский государственный университет (национальный исследовательский университет)». Россия, 454080 Челябинск, проспект Ленина, 76. E-mail: asiaevgd@susu.ru.

СОКОЛОВ Александр Николаевич, кандидат технических наук, доцент, заведующий кафедрой защиты информации ФГАОУ ВО «Южно-Уральский государственный университет (национальный исследовательский университет)». Россия, 454080 Челябинск, проспект Ленина, 76. E-mail: sokolovan@susu.ru.

ASYAEV Grigorii, Postgraduate Student, Department of Information Security, South Ural State University (National Research University). 76, Lenin prospekt, Chelyabinsk, Russia, 454080. E-mail: asiaevgd@susu.ru.

SOKOLOV Alexander, Ph.D., Associate Professor, Head of Department of Information Security, Federal State Autonomous Educational Institution of Higher Education “South Ural State University (national research university)”. 76, Lenin prospekt, Chelyabinsk, Russia, 454080. E-mail: sokolovan@susu.ru.

Материалы к публикации отправлять по адресу E-mail: urvest@mail.ru
в редакцию журнала «Вестник УрФО. Безопасность в информационной сфере».

Или по почте по адресу: Россия, 454080, г. Челябинск, пр. им. Ленина, д. 76,
ЮУрГУ, Издательский центр.

ВЕСТНИК УрФО
Безопасность в информационной сфере № 1(35) / 2020
Подписано в печать 30.03.2020.
Дата выхода в свет 14.05.2020. Формат 70×108 1/16. Печать цифровая.
Усл.-печ. л. 7,35. Тираж 100 экз. Заказ 88/171.
Цена свободная.

Отпечатано в типографии Издательского центра ЮУрГУ.
454080, г. Челябинск, пр. им. В. И. Ленина, 76.

Bulletin of the Ural Federal District
Security in the Sphere of Information No. 1(35) / 2020
Signed to print March 30, 2020.
Date of publication of the 14.05.2020. Format 70×108 1/16. Screen printing.
Conventional printed sheet 7,35. Circulation – 100 issues. Order 88/171. Open price.

Printed in the printing house of the Publishing Center of SUSU.
76, Lenina Str., Chelyabinsk, 454080