ФГБОУ ВО «Пензенский государственный университет», г. Пенза
ФГКОУ ВО «Воронежский институт МВД России», г. Воронеж
ФАУ «ГНИИИ ПТЗИ ФСТЭК России», г. Воронеж
ФГБОУ ВО «Липецкий государственный педагогический университет», г. Липецк
ФГБОУ ВО «Рязанский радиотехнический университет», г. Рязань
ФГБОУ ВО «Оренбургский государственный университет», г. Оренбург
ФГБОУ ВО «МИРЭА – Российский технологический университет», г. Москва
ФГУП «18-й Центральный научно-исследовательский институт» МО РФ, г. Москва
ФГАНУ «Институт сверхвысокочастотной полупроводниковой электроники имени В. Г. Мокерова Российской академии наук», г. Москва
АО «Пензенский научно-исследовательский электротехнический институт», г. Пенза
Пензенский филиал АО «Научно-технический центр "Атлас"», г. Пенза
АО «Научно-производственное предприятие "Рубин"», г. Пенза
АО «Производственное объединение "Электроприбор"», г. Пенза
АО «Радиозавод», г. Пенза
АО «Системы управления», г. Москва
Общероссийская общественная организация «Российское научно-техническое общество радиотехники, электроники и связи имени А. С. Попова», г. Тула
«Научно-исследовательский и конструкторский институт радиоэлектронной техники» филиал ФГУП НПЦ «Производственное объединение "Старт"» имени М. В. Проценко, г. Заречный
ФГБОУ ВО «Петербургский государственный университет путей сообщения Императора Александра I», г. Санкт-Петербург
Филиал АО «ПНИЭИ» Научно-исследовательское предприятие «Аргус», г. Пенза
ООО «Научно-производственная фирма "Кристалл"», г. Пенза
Филиал ФГКВОУ ВО «Военная академия Ракетных войск стратегического назначения имени Петра Великого», г. Серпухов
Обособленное подразделение ОАО «ИнфоТеКС», г. Пенза
ООО «НПФ "КРУГ"», г. Пенза
ООО «Научно-производственное предприятие "БиоКрипт"», г. Пенза
ООО «АЛГОМАТ», г. Калининград

Безопасность информационных технологий

Сборник научных статей по материалам
IV Всероссийской научно-технической конференции
(г. Пенза, 3 июля 2022 г.)

В двух томах

Том 1

Пенза Издательство ПГУ 2022

УДК 681.322
Б39

Безопасность информационных технологий : сб. науч. ст. по материалам IV Всерос. науч.-техн. конф. (г. Пенза, 3 июля 2022 г.) : в 2 т. – Пенза : Изд-во ПГУ, 2022. – Т. 1. – 162 с.

Б39

ISBN 978-5-907666-13-9

Рассматриваются различные аспекты безопасности информационных технологий. Публикуемые материалы прошли рецензирование.

Издание предназначено для специалистов по безопасности информационных технологий, преподавателей, аспирантов, докторантов и студентов вузов.

УДК 681.322

URL: https://tsib.pnzgu.ru/BIT

Состав оргкомитета научно-технической конференции:

Председатель – Волчихин В. И., заслуженный деятель науки РФ, д.т.н., профессор, президент ФГБОУ ВО «Пензенский государственный университет» (г. Пенза).

Сопредседатель – Фунтиков В. А., к.т.н., генеральный директор АО «ПНИЭИ» (г. Пенза).

Авсентьев О. С., д.т.н., профессор ФГКОУ ВО «Воронежский институт МВД России» (г. Воронеж); **Безяев В. С.**, к.т.н., советник генерального директора АО «НПП "Рубин"» (г. Пенза); **Безяев А. В.**, к.т.н., ведущий научный сотрудник Пензенского филиала АО «НТЦ "Атлас"» (г. Пенза); **Боровский А. С.**, д.т.н., доцент, заведующий кафедрой управления и информатики в технических системах ФГБОУ ВО «Оренбургский государственный университет» (г. Оренбург); **Газин А. И.**, к.т.н., доцент кафедры информатики, информационных технологий и защиты информации ФГБОУ ВО «Липецкий государственный педагогический университет» (г. Липецк); **Гамкрелидзе С. А.**, д.т.н., профессор, директор ФГАНУ «Институт сверхвысокочастотной полупроводниковой электроники имени В. Г. Мокерова Российской академии наук» (ИСВЧПЭ РАН) (г. Москва); **Голов И. Ю.**, к.т.н., главный научный сотрудник ФГУП «18 ЦНИИ» МО РФ (г. Москва); **Грунтович М. М.**, к.ф-м.н., доцент, руководитель Обособленного подразделения ОАО «ИнфоТеКС» (г. Пенза); **Зефиров С. Л.**, к.т.н., доцент, заведующий кафедрой информационной безопасности систем и технологий ФГБОУ ВО «Пензенский государственный университет» (г. Пенза); **Егоров В. Ю.**, к.т.н., начальник I отделения АО «НТП "Криптософт"» (г. Пенза); **Егорова Н. А.**, д.т.н., доцент кафедры информационной безопасности систем и технологий ФГБОУ ВО «Пензенский государственный университет» (г. Пенза); **Иванов А. И.**, д.т.н., доцент, научный консультант АО «ПНИЭИ» (г. Пенза); **Иванов А. П.**, к.т.н., доцент, заведующий кафедрой технических средств информационной безопасности ФГБОУ ВО «Пензенский государственный университет» на базе АО «ПНИЭИ» (г. Пенза); **Иванов В. А.**, д.т.н., профессор, генеральный директор ООО «АЛГОМАТ» (г. Калининград); **Качалин С. В.**, к.т.н., заместитель начальника отделения АО «НПП "Рубин"» (г. Пенза); **Князьков В. С.**, д.т.н., профессор, главный научный сотрудник НИИ ФиПИ ФГБОУ ВО «Пензенский государственный университет» (г. Пенза); **Козлов Г. В.**, д.т.н., профессор, директор Политехнического института ФГБОУ ВО «Пензенский государственный университет» (г. Пенза); **Костров Б. В.**, д.т.н., профессор, заведующий кафедрой электронных вычислительных машин ФГБОУ ВО «Рязанский радиотехнический университет» (г. Рязань); **Кулагин В. П.**, д.т.н., профессор, заведующий кафедрой аппаратного, программного и математического обеспечения вычислительных систем Института комплексной безопасности и специального приборостроения ФГБОУ ВО «МИРЭА – Российский технологический университет» (г. Москва); **Лазарев В. М.**, д.т.н., профессор, руководитель Управления координации научно-технического развития АО «Системы управления» (г. Москва); **Малыгин А. Ю.**, д.т.н., профессор, директор научно-образовательного центра «Информационная безопасность систем и технологий» ФГБОУ ВО «Пензенский государственный университет» (г. Пенза); **Мамон Ю. И.**, д.т.н., доцент, председатель Тульской областной организации Общероссийской общественной организации «Российское научно-техническое общество радиотехники, электроники и связи имени А. С. Попова» (г. Тула); **Привалов А. А.**, д.воен.н., профессор, академик РАЕН, профессор ФГБОУ ВО «Петербургский государственный университет путей сообщения Императора Александра I» (г. Санкт-Петербург); **Пушкин В. А.**, к.т.н., доцент, заместитель директора НТЦ АО «Радиозавод» (г. Пенза); **Урядов Д. А.**, заместитель главного конструктора АО ФНПЦ «ПО "Старт" имени М. В. Проценко» (г. Заречный Пензенской обл.); **Финько О. А.**, д.т.н., профессор Краснодарского высшего военного училища имени генерала армии С. М. Штеменко (г. Краснодар); **Цибизов П. Н.**, к.т.н., доцент АО ФНПЦ «ПО "Старт" имени М. В. Проценко» (г. Заречный Пензенской обл.); **Цимбал В. А.**, д.т.н., профессор, заслуженный деятель науки РФ, профессор филиала ФГКВОУ ВО «Военная академия Ракетных войск стратегического назначения имени Петра Великого» (г. Серпухов); **Шехтман М. Б.**, к.т.н., председатель совета директоров ООО «НПФ "КРУГ"» (г. Пенза); **Шумкин С. Н.**, к.т.н., начальник управления ООО «НПФ "Кристалл"» (г. Пенза); **Язов Ю. К.**, д.т.н., профессор, главный научный сотрудник Управления ФАУ «ГНИИИ проблем технической защиты информации ФСТЭК России» (г. Воронеж).

Приказ
о подготовке и проведении Всероссийской научно-технической конференции «Безопасность информационных технологий» № 472/о от 27.05.2022

ISBN 978-5-907666-13-9

© Пензенский государственный университет, 2022

УДК 004.056

ПОВЫШЕНИЕ РОЛИ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ В СОВРЕМЕННЫХ УСЛОВИЯХ

В. И. Волчихин

Пензенский государственный университет, г. Пенза

Аннотация. Рассмотрены основные задачи, поставленные Президентом В. В. Путиным на заседании Госсовета, посвященного информационной безопасности государства.

Ключевые слова: информационная безопасность, кибератаки, биометрико-нейросетевая аутентификация, защита персональных данных

INCREASING THE ROLE OF INFORMATION SECURITY IN MODERN CONDITIONS

V. I. Volchikhin

Penza State University, Penza

Abstract. The main tasks set by President V. V. Putin at the meeting of the State Council devoted to the information security of the state are considered.

Keywords: information security, cyberattacks, biometric-neural network authentication, personal data protection

Активные процессы цифровизации и глобализации современного мира только повышают актуальность проблемы информационной безопасности государства. В связи с этими глобальными

вызовами президент Путин на заседании госсовета предложил обсудить вопросы защиты и надежной работы информационных систем и сетей связи, а также меры, призванные парировать внешние угрозы в этой сфере [1]. Это важнейший вопрос для суверенитета и безопасности, экономики и госуправления, для общественной стабильности, подчеркнул он. «Количество кибератак на российскую информационную инфраструктуру все последние годы постоянно растет... а с началом специальной военной операции на Донбассе, на Украине вызовы в этой сфере стали еще более острыми и серьезными, более масштабными», – «По сути, против России развязана настоящая агрессия, война в информационном пространстве» [1].

Президент отметил, что в разы увеличилось число кибератак, в том числе комплексных, – хакерам-одиночкам это не под силу. Атаки наносятся из разных стран, и при этом четко скоординированы. По сути, это действия государственных структур, подчеркнул президент, добавив, что в состав армий некоторых стран официально входят кибервойска [1].

В речи президента особо отмечено, что возросли целенаправленные попытки вывести из строя интернет-ресурсы объектов критической информационной инфраструктуры России. В первую очередь под ударом оказались СМИ, банки, массовые социально значимые порталы и сети [1].

Еще одним серьезным вызовом, инструментом санкционного давления на Россию Путин назвал ограничения на зарубежные информационные технологии, программы и продукты [1].

В развитие указа о дополнительных мерах по обеспечению информбезопасности президент предложил обсудить проект основ госполитики в области обеспечения безопасности критической информационной инфраструктуры.

Путин В. В. заметил, что на трети критически важных объектов, от которых зависит обороноспособность, стабильное развитие экономики и социальной сферы, нет подразделений по защите информации. Координация действий структур обеспечения информационной безопасности должна быть закреплена на стратегическом уровне, а ответственность за решение этих вопросов возложена на руководителей организаций [1].

Вторая задача – повышение защищенности информационных систем и сетей связи в государственных органах, большинство ресурсов уязвимы для массированных атак и деструктивного внешнего воздействия. Президент отметил, что принципиально важно свести на нет риски утечек конфиденциальной информации и персональных

данных и поручил продумать создание государственной системы защиты информации [1].

Третья задача – кардинальное снижение рисков использования зарубежных программ, вычислительной техники и телекоммуникационного оборудования. Процессы цифровизации должны быть максимально защищены от любого потенциального негативного воздействия извне, обозначил он. Для этого нужен переход на отечественную технику, технологии, программы и продукты [1]. Отмечено также, что с 2025 г. использование зарубежных средств защиты информации будет запрещено. Для укрепления технологического суверенитета нужно максимально быстро создать современную электронную компонентную базу, и президент рассчитывает, что результат будет уже скоро [1].

Актуальность докладов, включенных в программу работы конференции во многом совпадает с задачами, поставленными президентом страны В.В. Путиным, и свидетельствует о возрастании роли защиты информации, кибербезопасности [2], развития российской технологии биометрической аутентификации с использованием больших искусственных нейронных сетей [3], ее применение по защите персональных данных пользователей [4], а также продолжение ранее начатых исследований по разработке и внедрению новых типов искусственных нейронов в нейросетевых преобразователя биометрия-код [5–8], позволяющих усилить стойкость средств аутентификации личности к атакам подбора кода ключа доступа.

Отмечу, что IV Всероссийская научно-техническая конференция «Безопасность информационных технологий» также проходит в смешанном режиме: очно и дистанционно, на этот счет имеется определенный положительный опыт проведения предыдущей конференции в таком формате.

Хочу пожелать всем участникам конференции хорошей работы, крепкого здоровья и новых творческих успехов.

Список литературы

1. Российская газета. URL: https://rg.ru/2022/05/22/vladimir-putin-obsudil-novuiu-politiku-informacionnoj-bezopasnosti.html

2. Костарев С. В., Карганов В. В., Липатников В. А. Технологии защиты информации в условиях кибернетического противоборства : монография. СПб. : ВАС, 2020. 716 с.

3. Волчихин В. И., Иванов А. И., Фунтиков В. А., Малыгина Е. А. Перспективы использования искусственных нейронных сетей с много-

уровневыми квантователями в технологии биометрико-нейросетевой аутентификации // Известия высших учебных заведений. Поволжский регион. Технические науки. 2013. № 4 (28). С. 88–99.

4. Язов Ю. К., Волчихин В. И., Иванов А. И., Фунтиков В. А., Назаров И. Г. Нейросетевая защита персональных биометрических данных / под ред. Ю. К. Язова. М. : Радиотехника, 2012. 157 с.

5. Волчихин В. И., Иванов А. И., Перфилов К. А., Малыгина Е. А., Серикова Ю. И. Быстрый алгоритм обучения больших сетей искусственных нейронов квадрата среднего геометрического плотностей распределения значений многомерных биометрических данных // Известия высших учебных заведений. Поволжский регион. Технические науки. 2018. № 3 (47). С. 38–45.

6. Волчихин В. И., Иванов А. И., Малыгина Е. А., Серикова Ю. И. Сопоставление мощностей двух типов искусственных нейронов, осуществляющих обогащение биометрических данных в линейном и квадратичном пространствах //Известия высших учебных заведений. Поволжский регион. Технические науки. 2018. № 3 (47). С. 59–66.

7. Волчихин В. И., Иванов А. И., Вятчанин С. Е., Малыгина Е. А. Абсолютно устойчивый алгоритм автоматического обучения сетей вероятностных нейронов «Крамера – фон Мизеса» на малых выборках биометрических данных // Известия высших учебных заведений. Поволжский регион. Технические науки. 2017. № 2 (42). С. 55–65.

8. Малыгина Е. А. Биометрико-нейросетевая аутентификация: перспективы применения сетей квадратичных нейронов с многоуровневым квантованием биометрических данных : препринт. Пенза : Изд-во ПГУ, 2020. 110 с.

Для цитирования: Волчихин В. И. Повышение роли информационной безопасности в современных условиях // Безопасность информационных технологий : сб. науч. ст. по материалам IV Всерос. науч.-техн. конф. : в 2 т. Пенза : Изд-во ПГУ, 2022. Т. 1. С. 3–6.

УДК 004.056

ОЦЕНКА БИТОВОГО ПОТОКА МЕЖДУ ДОВЕРЕННЫМ ПРОЦЕССОРОМ SIM-КАРТЫ И НЕДОВЕРЕННЫМ ОКРУЖЕНИЕМ ПРИ НЕЙРОСЕТЕВЫХ ВЫЧИСЛЕНИЯХ В РЕЖИМЕ, ЗАЩИЩЕННОМ ГОМОМОРФНЫМ ШИФРОВАНИЕМ

В. С. Князьков[1], А. И. Иванов[2], К. Н. Савинов[3]

[1,3]Пензенский государственный университет, г. Пенза
[2]Пензенский научно-исследовательский электротехнический институт, г. Пенза

Аннотация. Показано, что при использовании для доверенных вычислений контроллера SIM-карты с ограниченными ресурсами требуется привлечение внешних ресурсов для выполнения биометрико-нейросетевых преобразований, защищенных гомоморфным шифрованием. Оцениваются битовые потоки последовательного интерфейса между SIM-картой и внешней средой для трех типов сетей искусственных нейронов: многоуровневых персептронов, многоуровневых квадратичных нейронов и нейронов среднего геометрического. Сделан вывод о возможности создания специальных модификаций искусственных нейронов, ориентированных на снижение битового потока между SIM-картой и внешней средой, выполняющей вычисления, защищенные гомоморфным шифрованием.

Ключевые слова: программирование в остаточных классах, искусственные нейроны, гомоморфное шифрование, доверенные вычисления

EVALUATION OF A BITFLOW BETWEEN A TRUSTED PROCESSOR OF A SIM-CARD AND A UNTRUSTED ENVIRONMENT DURING NEURAL NETWORK COMPUTATIONS IN A MODE PROTECTED BY HOMOMORPHY ENCRYPTION

V. S. Knyazkov[1], A. I. Ivanov[2], K. N. Savinov[3]

[1,3]Penza State University, Penza
[2]Penza Research Electrotechnical Institute, Penza

Abstract. It is shown that when a SIM-card controller with limited resources is used for trusted computing, external resources are required to perform biometric-neural network transformations protected by homomorphic encryption. The bit

© Князьков В. С., Иванов А. И., Савинов К. Н., 2022

streams of the serial interface between the SIM card and the external environment are estimated for three types of artificial neuron networks: multilevel perceptron's, multilevel quadratic neurons, and geometric mean neurons. It is concluded that it is possible to create special modifications of artificial neurons aimed at reducing the bit stream between the SIM card and the external environment that performs calculations protected by homomorphic encryption.

Keywords: programming in residual classes, artificial neurons, homomorphic encryption, trusted computing

Доверенные операции криптографии и нейросетевой биометрии для мобильных пользователей должны выполняться в массовых процессорах низкой стоимости, например, SIM-карт (на сегодняшний день это типовое техническое решение). Пример организации такого процессора дан на рис. 1.

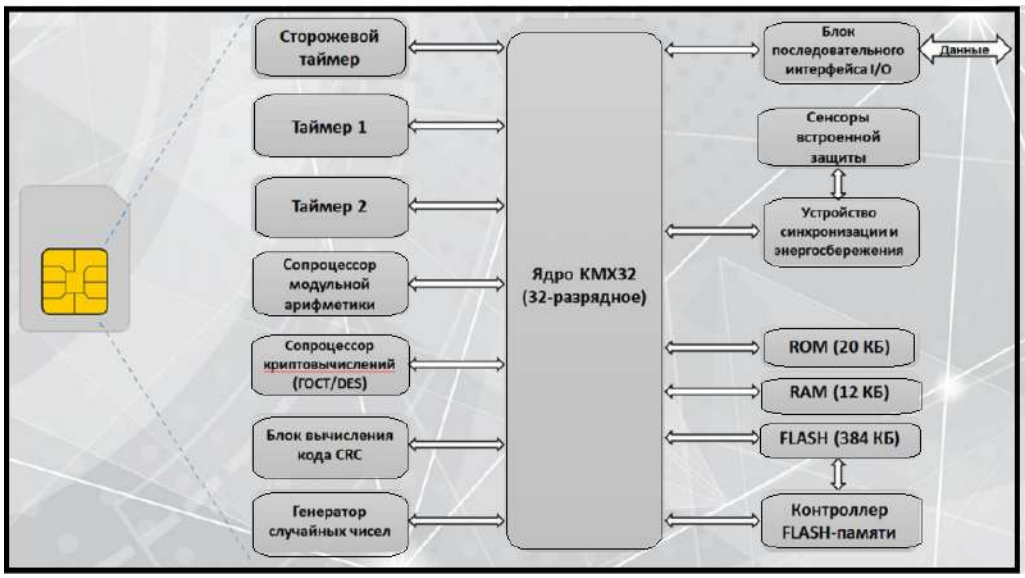

Рис. 1. Структура массового доверенного вычислителя SIM-карты мобильного устройства

К сожалению, процессор SIM-карт имеют ограниченную память, ограниченные по потреблению энергии (по производительности), т.е. размещать в одной SIM-карт криптографию удается, а дополнительно размещать нейросетевую биометрию достаточно сложно. В связи с этим приходится доверенные вычислительные ресурсы занимать у внешней не доверенной вычислительной среды через внешний последовательный интерфейс (верхний правый угол рис. 1). При этом необходимо обеспечить достаточный поток информации через внешний последовательный интерфейс и так же необходимо обеспечить гомоморфное шифрование [1] для внешней реализации элементов нейросетевого решающего правила [2–4].

При реализации, рассматриваемой схемы вычислений, необходимо выполнить гомоморфное шифрование биометрических данных почти в реальном времени и одновременно выполнять гомоморфное расшифровывание получаемых обратно после внешних вычислений данных. Соответствующая блок-схема вычислений представлена на рис. 2.

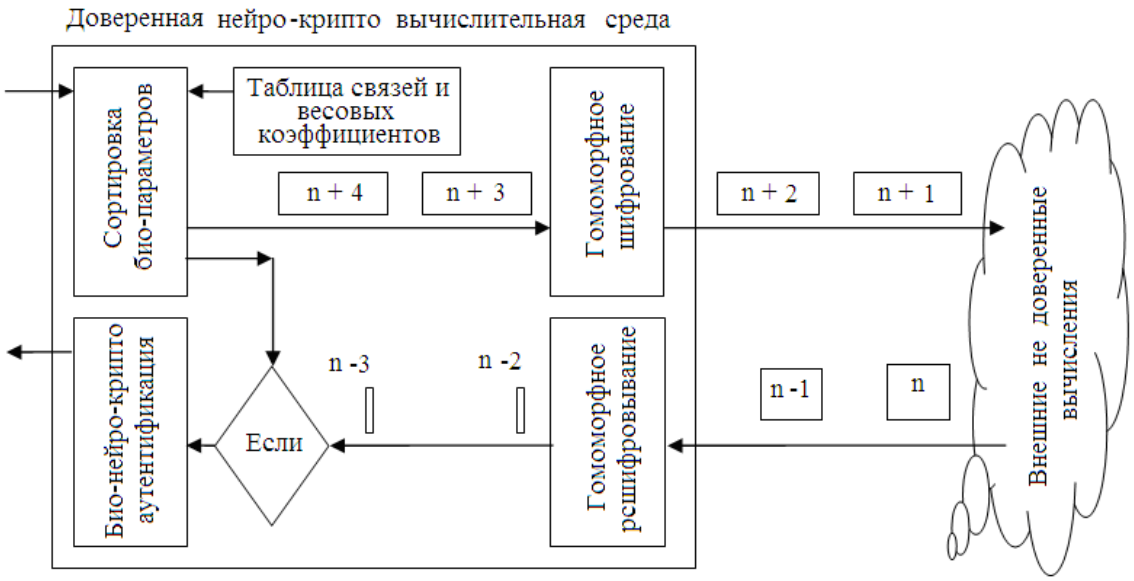

Рис. 2. Доверенные вычисления, ориентированные на привлечение внешнего не доверенного вычислителя при обычном гомоморфном шифровании (гомоморфизм только относительно двух типов операций «сложения» и «умножения»)

Следует отметить, что внешний вычислитель имеет неограниченный ресурс мощности и памяти, модель нейросетевого решающего правила может быть создана за ранее и заранее гомоморфно зашифрована. К сожалению, стандартизованное решение гомоморфного шифрования [1] не способно работать с очень большими нейросетевыми решающими правилами. Обычное шифрование работоспособно для защиты текстов любой длины, после шифрования она расшифровываются без особых проблем. Гомоморфное шифрование нейросетевых решающих правил не может быть слишком сложным, так как сложные гомоморфные вычисления при их применении накапливают ошибки. В связи с этим сложное нейросетевое решающее правило должно быть разбито на множество мелких решающих правил, например, реализующих только один нейрон. В этом случае ошибки не накапливаются, более того, ошибки могут устраняться при нейросетевых вычислениях.

В связи с тем, что рынок средств биометрической аутентификации регулируется «ФСТЭК России» и «ФСБ России», желательно использовать только рекомендуемые регуляторами технические решения. На данный момент в России могут быть использованы сети персептронов, автоматически обученные по ГОСТ Р 52633.5–2011. Этот тип нейросетевых преобразователей должен быть ориентирован на использование 256 искусственных нейронов с бинарными выходными квантователями. В этом случае нейросеть должна сопрягаться с отечественными криптографическими алгоритмами шифрования, цифровой подписи или аутентификации с длинного криптографического ключа в 256 бит.

При этом каждый отдельный нейрон будет описываться связями, приведенными ниже:

$$\begin{cases} y \leftarrow \sum_{i=1}^{16} a_i \cdot x_i + c \\ z(y) \leftarrow "0" \ \ if \ \ y \le 0 \\ z(y) \leftarrow "1" \ \ if \ \ \ y > 0 \\ P_1 \approx 0,01 \\ P_2 \approx 0.5 \end{cases}, \qquad (1)$$

где x_i – входные биометрические данные нейрона; a_i – весовые коэффициенты нейрона; P_1 – вероятность ошибок первого рода (ошибочный отказ в доступе); P_2 – вероятность ошибок второго рода (ошибочный пропуск «Чужого»).

Для того, чтобы оценить поток, через выход информационного обмена «узкого горлышка» SIM-карты, необходимо задаться конкретной схемой реализации вариантов гомоморфного шифрования [5]. Будем исходить из того, что 16 биометрических параметров и 16 весовых коэффициентов выражения (1) при открытом варианте реализации нейросетевого решающего правила кодируются 8-ми битными бинарными кодами. В этом случае выходной битовый поток из процессора SIM-карты составит 16 байт. После внешнего вычисления состояния сумматора нейрона (1) обратно в SIM-карту придется возвращать 2 байта.

Будем исходить из того, что схема гомоморфного шифрования выполнена в модулярной арифметике [5], опираясь на разложение элементов решающего правила по пяти модулям $\{q_1, q_2, q_3, q_4, q_5,\}$. В этом случае для перехода к модулярному представлению данных гомоморфное шифрование придется выполнять, вычисляя пять остатков по разным модулям:

$$\begin{cases} r_{1,i} \leftarrow \left(x_i - round\left\{x_i / q_1, 0\right\}\right) \\ r_{2,i} \leftarrow \left(x_i - round\left\{x_i / q_2, 0\right\}\right) \\ r_{3,i} \leftarrow \left(x_i - round\left\{x_i / q_3, 0\right\}\right). \\ r_{4,i} \leftarrow \left(x_i - round\left\{x_i / q_4, 0\right\}\right) \\ r_{5,i} \leftarrow \left(x_i - round\left\{x_i / q_5, 0\right\}\right) \end{cases} \quad (2)$$

Если исходить из того, что каждый остаток по последовательному интерфейсу SIM-карты будет передаваться 8-ми битными числами, то гомоморфно зашифрованный поток биометрических данных для одного нейрона составит 16*5 = 80 байт. Соответственно 256 нейронов дадут выходной поток 256*16*5 = 20480 байт. Обратный информационный поток в SIM-карту после внешних доверенных вычислений будет в пять раз меньше 256*16 = 4096 байт.

При реализации нейросетевых преобразований в доверенной вычислительной среде крайне важны особенности используемых решающих правил. Практика применения сетей персептронов, обученных по ГОСТ Р 52633.5–2011 показала, что примерно 1/3 часть биометрических данных используется не рационально. Более рациональное использование входных данных приводит к тому, что практически то же самое линейное накопление данных (1) позволяет использовать троичные квантователи с двумя порогами:

$$\begin{cases} y \leftarrow \sum_{i=1}^{16} a_i \cdot x_i + c \\ z(y) \leftarrow "00" \ \ if \ \ y \le k_1 \\ z(y) \leftarrow 01 \ \ if \ \ k_1 < y \le k_2 \\ z(y) \leftarrow "10" \ \ if \ \ y > k_2 \\ P_1 \approx 0{,}01 \\ P_2 \approx 0{,}333 \end{cases}, \quad (3)$$

Последнее означает, что каждый троичный нейрон отвечает за пару бит выходного кода нейросети. То есть для получения 256 битного ключа требуется всего 128 искусственных нейронов. Это эквивалентно двухкратному снижению байтового потока от SIM-карты во внешнюю вычислительную среду 128*16*5 = 10 240 байт.

Еще большего сокращения потока от SIM-карты во внешнюю вычислительную среду удается добиться, если перейти к квадратичным нейронам с четверичным выходным квантователем:

$$\begin{cases} y \leftarrow \sum_{i=1}^{8} \{a_i \cdot x_i - c_i\}^2 \\ z(y) \leftarrow "00" \ \ if \ \ y \le k_1 \\ z(y) \leftarrow "01" \ \ if \ \ k_1 < y \le k_2 \\ z(y) \leftarrow "10" \ \ if \ \ k_2 < y < k_3 \\ z(y) \leftarrow "11" \ \ if \ \ y > k_3 \\ P_1 \approx 0,01, \quad P_2 \approx 0,25 \end{cases}. \tag{4}$$

Из-за того, что эллиптические решающие правила (4) лучше выделяют образ «Свой» в сравнении с линейными гиперплоскостями (1), удается снизить число входов у нейроннов с 16 до 8. Это приводит к сокращению потока от SIM-карты во внешнюю вычислительную среду еще в два раза 128*8*5 = 5120 байт.

Таким образом, выходной и обратный потоки SIM-карты сильно зависят от типа, используемых решающим правилом искусственных нейронов. На текущий момент мы имеем десятки искусственных нейронов с разными выходными квантователями [6–8]. Видимо под учет тех или иных особенностей гомоморфного шифрования придется модифицировать те или иные искусственные нейроны.

Список литературы

1. ISO/IEC 18033-6: 2019 IT Security techniques-Encryption algorithms – Pard 6: Homomorphic encryption.

2. Князьков В. С., Иванов А. И., Безяев А. В. Необходимость расширения функциональных возможностей гомоморфного шифрования для защиты нейросетевых решающих правил биометрических приложений искусственного интеллекта // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. Пенза : Изд-во ПГУ, 2020. С. 5–10.

3. Князьков В. С., Иванов А. И., Безяев А. В., Лукин В. С. Бескомпроматное привлечение сторонних ресурсов низкого доверия для выполнения вычислений высокого доверия в SIM картах и микро SD-картах с защитой персональных биометрических данных нейро-гомоморфным шифрованием // Безопасность информационных технологий : сб. науч. ст. по материалам III Всерос. науч.-техн. конф. Пенза : Изд-во ПГУ, 2020. С. 55–62.

4. Иванов А. И., Князьков В. С. Перспектива многократного увеличения ресурсов доверенных вычислений за счет привлечения гибрида нейросетевой обработки биометрии и гомоморфного шифрования // «Состояние и перспективы развития современной науки по направлению «Техническое зрение, распознавание образов» : материалы III Всерос. науч.-техн. конф. (г. Анапа, 18 марта 2021). Анапа : Военный инновационный технополис «ЭРА», 2021. С. 173–176.

5. Червяков Н. И., Сахнюк П. А., Шапошников А. В., Махота А. Н. Нейрокомпьютеры в остаточных классах : учеб. пособие. М. : Радиотехника, 2003. 272 с.

6. Малыгина Е. А. Биометрико-нейросетевая аутентификация: перспективы применения сетей квадратичных нейронов с многоуровневым квантованием биометрических данных : препринт. Пенза : Изд-во ПГУ, 2020. 114 с. ISBN 978-5-907262-88-1

7. Иванов А. И., Банных А. Г., Куприянов Е. Н., Лукин В. С., Перфилов К. А., Савинов К. Н. Коллекция искусственных нейронов эквивалентных статистическим критериям для их совместного применения при проверке гипотезы нормальности малых выборок биометрических данных // Безопасность информационных технологий : сб. науч. ст. по материалам I Всерос. науч.-техн. конф. (г. Пенза, 24 апреля 2019 г.). Пенза, 2019. С. 156–164.

8. Иванов А. И., Куприянов Е. Н. Защита искусственного интеллекта : ортогонализация статистико-нейросетевого анализа малых выборок биометрических данных : препринт. Пенза : Изд-во ПГУ, 2020. 72 с.

Для цитирования: Князьков В. С., Иванов А. И., Савинов К. Н. Оценка битового потока между доверенным процессором SIM-карты и недоверенным окружением при нейросетевых вычислениях в режиме, защищенном гомоморфным шифрованием // Безопасность информационных технологий : сб. науч. ст. по материалам IV Всерос. науч.-техн. конф. : в 2 т. Пенза : Изд-во ПГУ, 2022. Т. 1. С. 7–13.

УДК 004.032.26;004.827

ПРОСТОЙ АЛГОРИТМ ИТЕРАЦИОННОГО ПОДБОРА РАЗМАХА СЛУЧАЙНЫХ АДДИТИВНЫХ МУТАЦИЙ ПРИ РАЗМНОЖЕНИИ ДАННЫХ ПРИМЕРОВ ОБРАЗА «ЧУЖОЙ»

А. Г. Банных[1], А. П. Иванов[2], С. В. Туреев[3]

[1,2]Пензенский государственный университет, г. Пенза
[3]АО «Концерн "Созвездие"», г. Воронеж

Аннотация. Предложен простой алгоритм выбора амплитуды белого шума, аддитивно подмешиваемого при тестировании нейросетевых преобразователей биометрических данных в длинный код аутентификации.

Ключевые слова: информационная безопасность, тестирование, нейронные сети, преобразование биометрии в код аутентификации, расстояния Хэмминга

SIMPLE ALGORITHM FOR ITERATIVE SELECTION OF RANDOM ADDITIVE MUTATIONS RANGE DURING REPLACEMENT OF THESE EXAMPLES OF THE «ALIEN» IMAGE

A. G Bannykh[1], A. P. Ivanov[2], S. V. Tureev[3]

[1,2]Penza State University, Penza
[3]Concern «Sozvezdie», Voronezh

Abstract. A simple algorithm is proposed for selecting the amplitude of white noise, which is additively mixed in when testing neural network converters of biometric data into a long authentication code.

Keywords: information security, testing, neural networks, conversion of biometrics into an authentication code, Hamming distances

Введение

В случае формирования тестовой базы биометрических образов «Чужой» по требованиям ГОСТ Р 52633.1 [1] для каждого образа сохраняют 20 примеров. В этом случае мы можем воспользоваться рекомендациями ГОСТ Р 52633.2 [2] и получить достаточно большое число промежуточных биометрических образов, скрещиванием

© Банных А. Г., Иванов А. П., Туреев С. В., 2022

образов-родителей. По стандарту ГОСТ Р 52633.2 [2] существует еще один подход к размножению тестовых образов за счет применения случайных мутаций к данным образов-родителей перед их скрещиванием. Одним из основных недостатков второго подхода является то, что стандарт не содержит четких рекомендаций по выбору амплитуды случайных мутаций, используемых при размножении данных.

Целью данной работы является устранение неопределенности по выбору амплитуды (стандартного отклонения), используемых случайных чисел (мутаций). Необходимо объединить процедуры морфинг-скрещивания данных с процедурами аддитивного добавления случайных мутаций.

Дополнительная сортировка скрещиваемых данных

Как основу соединения разных процедур воспользуемся процедурой упорядочивания данных скрещиваемых примеров «Чужой» в пространстве расстояний Хэмминга. Очевидным является то, что примеры одного образа «Чужой-*k*» будут иметь отклики нейросети с разными значениями кодов. В свою очередь каждый код будет иметь свое расстояние Хэмминга по отношению к коду образа «Свой». Эта ситуация отображена на рис. 1.

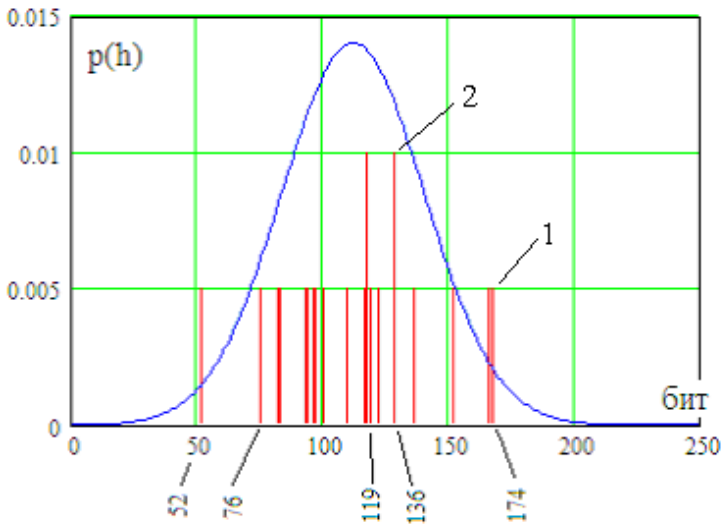

Рис. 1. Пример распределения расстояний Хэмминга
для 20 примеров образа «Чужой»

Из рис. 1 мы видим, что 20 примеров образа «Чужой» дает последовательность расстояний Хэмминга от 52 до 174 бит. При этом наиболее близким примером образа «Чужой» к образу «Свой» будет являться пример с минимальным значением расстояния Хэмминга $\min(h) = h_1 = 52$ бита. Наиболее далеким будет являться пример с максимальным значением расстояния Хэмминга $\max(h) = h_{20} = 174$ бита.

Все остальные примеры, рассматриваемого образ «Чужой» так же могут быть упорядочены между собой $h_1 = 52 \le h_2 = 76 \le h_3 \le \ldots \le h_{20} = 174$ по их расстоянию Хемминга до кода образа «Свой».

Итерационный подбор границ размаха случайных мутаций для ближайшего примера образа «Чужой»

Будем исходить из того, что случайные мутации могут существенно значимыми и мало значимыми. Если как источник мутаций мы будем использовать программный генератор с нулевым математическим ожиданием $E(x) = 0$ и некоторым стандартным отклонением $\sigma(x)$. Естественно, что программный генератор с нулевым стандартным отклонением $\sigma(x) = 0$, должен давать мало значимые мутации (никак не влияющие на итоговый результат, расстояние Хэмминга не меняется). Однако, если мы увеличим стандартное отклонение до некоторой величины $\sigma_{min}(x) < 0$, то сможем наблюдать рост расстояния Хэмминга с 52 бит до 53 бит.

Очевидным так же является монотонный рост расстояний Хэмминга при монотонном увеличении стандартного отклонения $\sigma(x)$. Так же очевидной является ситуация, когда при некотором значении $\sigma_{max}(x) << \sigma_{min}(x)$ должны появляться расстояния Хэмминга $h_{1,M} = 76 = h_2$. Дальнейшее увеличение стандартного отклонения программного генератора мутаций не имеет смысла, так как неограниченный рост стандартного отклонения генератора мутаций приводит к полной утрате внутренних корреляционных связей синтезируемых биометрических образов [3]. В место не оправданной утраты внутренних корреляционных связей синтезируемых биометрических данных более целесообразным является переход к использованию для размножения мутаций следующего, второго примера образа «Чужой» с исходным расстоянием Хэмминга $h_2 = 76$.

Повторяя, описанные выше процедуры подбора значений стандартного отклонения программного генератора мутаций, мы имеем возможность ограничить с верху подбираемые значения стандартных отклонений. В конечном итоге это позволяет объединить процедуры синтеза промежуточных биометрических образов скрещиванием их примеров с аналогичными процедурами размножения данных мутациями. Рассматриваемые модификации показали свою высокую эффективность при экономичном табличном вычислении энтропии биометрических образов «Чужой» [4, 5] мало потребляющими доверенными процессорами SIM-карт, микро SD-карт, RFID идентификационных карт.

Предположительно, изложенные в данной работе подходы могут быть положены в основу разработки модифицированной версии

национального стандарта взамен действующего в настоящее время отечественного стандарта ГОСТ Р 52633.2–2010.

Список литературы

1. ГОСТ Р 52633.1–2009. Защита информации. Техника защиты информации. Требования к формированию баз естественных биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации.
2. ГОСТ Р 52633.2–2010. Защита информации. Техника защиты информации. Требования к формированию синтетических биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации.
3. Туреев С. В., Малыгина Е. А., Солопов А. И. Методика формирования тестовых баз для проверки качества обучения нейросетевых преобразователей биометрия-код // Безопасность информационных технологий : сб. науч. ст. по материалам I Всерос. науч.-техн. конф. (г. Пенза, 24 апреля 2019 г.). Пенза, 2019. С. 90–102.
4. Волчихин В. И., Иванов А. И., Банных А. Г. Регуляризация вычисления энтропии выходных состояний нейросетевого преобразователя биометрия-код, построенная на размножении малой выборки исходных данных // Известия высших учебных заведений. Поволжский регион. Технические науки. 2017. № 4. С. 14–23.
5. Свидетельство о государственной регистрации программы для ЭВМ № 2021614767. Калькулятор для вычисления энтропии кодов 256 бит на малых выборках / А. Г. Банных, А. И. Иванов, А. П. Иванов, А. А. Пирогов. № 2021613966 ; заявл. 26.03.2021 ; зарег. 30.03.2021.

Для цитирования: Банных А. Г., Иванов А. П., Туреев С. В. Простой алгоритм итерационного подбора размаха случайных аддитивных мутаций при размножении данных примеров образа «Чужой» // Безопасность информационных технологий : сб. науч. ст. по материалам IV Всерос. науч.-техн. конф. : в 2 т. Пенза : Изд-во ПГУ, 2022. Т. 1. С. 14–17.

УДК 519.24; 57.017

ИЕРАРХИЧЕСКАЯ СТРУКТУРА СВЯЗЕЙ САМОКОРРЕКТИРУЮЩИХСЯ КОДОВ, ОРИЕНТИРОВАННЫХ НА НЕЙРОСЕТЕВОЕ ОБОБЩЕНИЕ МНОЖЕСТВА СТАТИСТИЧЕСКИХ КРИТЕРИЕВ ПРОВЕРКИ ГИПОТЕЗЫ НЕЗАВИСИМОСТИ МАЛЫХ ВЫБОРОК

Т. А. Золотарева[1], А. В. Безяев[2], Ю. И. Олейник[3]

[1]Липецкий государственный педагогический университет имени П. П. Семенова-Тян-Шанского, г. Липецк
[2]Пензенский филиал научно-технического центра «Атлас», г. Пенза
[3]Радиозавод, г. Пенза

Аннотация. Рассматривается переход от самокорректирующихся кодов с высокой избыточностью и одноуровневой иерархией к более эффективным кодам с двухуровневой иерархией. Дана схема объединения в группы по три параметра, в каждой группе корректировка кодов выполняется голосованием по большинству состояний.

Ключевые слова: кодовые конструкции, способные обнаруживать и исправлять ошибки; гипотеза независимости малых выборок, искусственные нейроны

HIERARCHICAL STRUCTURE OF RELATIONSHIPS OF SELF-CORRECTING CODES ORIENTED ON THE NEURAL NETWORK GENERALIZATION OF A SET OF STATISTICAL CRITERIA FOR VERIFICATION OF THE HYPOTHESIS OF INDEPENDENCE OF SMALL SAMPLES

T. A. Zolotareva[1], A. V. Bezyaev[2], Yu. I. Oleynik[3]

[1]Lipetsk State Pedagogical University named after P. P. Semenov-Tyan-Shansky, Lipetsk
[2]Penza branch of Scientific and Technical Center «Atlas», Penza
[3]Radiozavod, Penza

Abstract. The transition from self-correcting codes with high redundancy and a one-level hierarchy to more efficient codes with a two-level hierarchy is considered. A scheme is given for grouping into groups of three parameters; in each group, the codes are corrected by voting on the majority of states.

Keywords: code constructs capable of detecting and correcting errors; small sample independence hypothesis, artificial neurons

© Золотарева Т. А., Безяев А. В., Олейник Ю. И., 2022

Введение

При решении ряда практических задач (медицины, биологии, биометрии, экономики, физики, химии, …) трудно получать выборки данных большого объема. Как следствие приходится статистически анализировать малые выборки объемом от 16 до 21 опытов. Обычно при статистическом анализе оценивают первые статистические моменты, такие как: математическое ожидание, стандартное отклонение и коэффициент корреляции.

Статистическому анализу малых выборок уделялось значительное внимание, как в прошлом веке, так и в XXI веке. Достигнутые результаты в этом направлении, отражены в справочнике по математической статистике [1]. Справочник содержит описание порядка 200 статистических критериев, в том числе этот источник содержит описание более 30-ти статистических критериев, созданных для проверки гипотезы независимости данных.

Проблему проверки гипотезы независимости данных иллюстрирует рис. 1, где отображены распределения значений коэффициентов корреляции, вычисленные по классической формуле Пирсона-Эджуорта-Эудлона (1890 г.) [2].

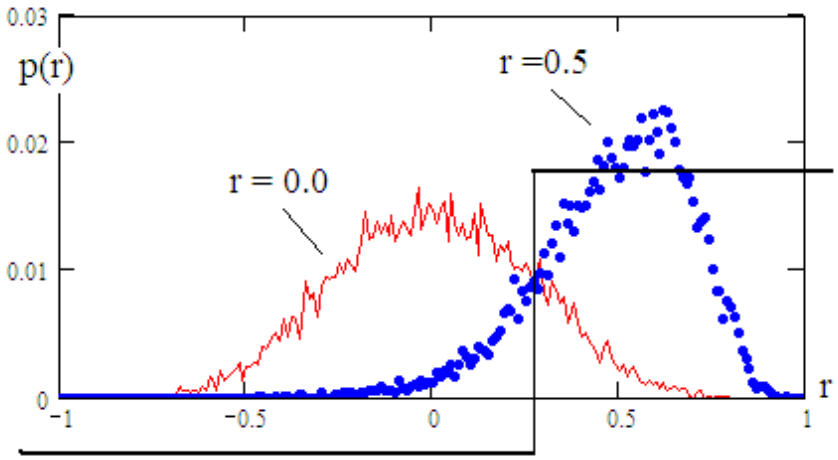

Рис. 1. Распределение значений коэффициентов корреляции, вычисленных по классической формуле для малой выборки из 16 опытов

Из-за малого объема выборки в 16 опытов независимые данные $r = 0{,}0$ дают значения, попадающие в интервал от $r = -0{,}75$ до $r = +0{,}75$. К сожалению, классическая формула вычисления коэффициентов корреляции построена на использовании двух математических ожиданий $-E(.)$ и двух стандартных отклонений $-\sigma(.)$:

$$r(x,y)=\frac{1}{16}\sum_{i=1}^{16}\frac{\left(E(x)-x_i\right)\cdot\left(E(y)-y_i\right)}{\sigma(x)\cdot\sigma(y)}. \qquad (1)$$

Вычисление младших статистических моментов на малых выборках приводит к появлению значительных ошибок $\Delta E(.)$ и $\Delta\sigma(.)$. При вычислениях вида (1) ошибки исходных данных $\Delta E(.)$ и $\Delta\sigma(.)$ накапливаются. Это одна из основных причин, по которой оценка коэффициентов корреляции по классической формуле (1), имеет значительную ошибку – $\Delta r(x, y)$. Эта ошибка оказывается приемлемой только для достаточно больших выборок объемом от 160 до 200 примеров.

Нейросетевое обобщение множества статистических критериев проверки гипотезы независимости малых выборок биометрических данных

Особо остро стоит проблема статистических оценок на малых выборках в нейросетевой биометрии. Так национальный стандарт ГОСТ Р 52633.5–2001 [10] ориентирован на автоматическое обучение сети искусственных нейронов на 20 примерах образа «Свой». При этом по большинству из биометрических параметров «хорошей» обучающей выборки законы распределения должны быть близки к нормальным. То есть мы должны перед обучением проверять гипотезу нормальности, пользуясь одним из 21 известных статистических критериев проверки гипотезы нормальности.

Биометрико-нейросетевые технологии аутентификации ориентированы на рынок средств защиты информации, регулируемый двумя государственными службами «ФСБ России» и «ФСТЭК России». То есть массовые продукты биометрико-нейросетевой аутентификации должны будут иметь сертификаты этих двух организаций. В связи с этим «ФСТЭК России» под будущую сертификацию создал 7 действующих стандартов усилиями технического комитета по стандартизации № 362 «Защита информации». Технический комитет по стандартизации № 26 «Криптографическая защита информации» разработал техническую спецификацию по рекуррентному шифрованию и расшифровыванию таблиц данных обученных нейронов. Ввод в действие на территории РФ этой технической спецификации предположительно произойдет во второй половине 2021 года.

Следует отметить, что разработанные отечественные стандарты и техническая спецификация распространяются только на нейронные сети с накоплением данных в линейном пространстве. Можно говорить о высоком уровне национальной стандартизации технологии

обработки биометрических данных сетями персептронов (нейронов, выполняющих накопление данных взвешенным суммированием).

Видимо, следующим классом актуальным для стандартизации будут сети, состоящие из квадратичных нейронов. Этот тип искусственных нейронов осуществляет накопление относительно бедных «сырых» биометрических данных в квадратичных пространствах.

В общем случае все квадратичные нейроны могут быть описаны через квадратичную форму:

$$\begin{cases} e^2 \leftarrow \left[E(x)-\overline{x}\right]^T \cdot [r]^{-1} \cdot \left[E(x)-\overline{x}\right] \\ z\left(e^2\right) \leftarrow "00" \ \textit{если} \rightleftarrows \quad e^2 < k_1 \\ z\left(e^2\right) \leftarrow "01" \ \textit{если}\ k_1 < e^2 < k_2 \\ z\left(e^2\right) \leftarrow "10" \ \textit{если}\ k_2 < e^2 < k_3 \\ z\left(e^2\right) \leftarrow "11" \ \textit{если} \quad e^2 > k_3 \end{cases} . \qquad (2)$$

где $\overline{x}$ – вектор «сырых» нормированных биометрических параметров с единичным стандартным отклонением $\sigma(x) = 1$; $[r]^{-1}$ – обратная корреляционная матрица «сырых» биометрических параметров; z(.) – четырех уровневый выходной квантователь искусственного нейрона, имеющий три порога сравнения – $\{k_1, k_2, k_3\}$.

Самой сложной операцией в настройке искусственного нейрона (2) является плохо обусловленная операция вычисления обратной корреляционной матрицы. Очевидно, что эта операция становится устойчивой, если входные данные не коррелированы (не зависимы). В этом случае корреляционная матрица оказывается единичной и ее обращение становится устойчивым.

Тем не менее, проблема точного вычисления коэффициентов корреляции на малых выборках остается. Так если мы получаем данные объемом в 16 примеров, то мы не способны с приемлемой точностью проверять гипотезу независимости $r = 0{,}0$. Классическая формула (1) в место реального значения $r = 0{,}0$ будет давать значения лежащие в интервале от $r = -0{,}75$ до $r = +0{,}75$ (см. рис. 1). Столь значительные отклонения неприемлемы для практики, однако мы все же можем создать на базе классической формулы (1) искусственный нейрон, способный различать состояние $r = 0{,}0$ и состояние $r = 0{,}5$ с вероятностями ошибок первого и второго рода $P_1 \approx P_2 \approx 0{,}146$.

Очевидным является то, что для всех 15 известных классических статистических критериев проверки гипотезы независимости [1]

может быть построен свой искусственный нейрон. То есть уже сегодня, может быть, построена сеть из 15 искусственных нейронов, обобщающая, созданные в прошлом веке статистические критерии проверки гипотезы независимости.

По аналогии с критериями проверки гипотезы проверки нормальности данных были проведены работы по созданию новых статистических критериев проверки гипотезы независимости. В частности, был создан критерий (искусственный нейрон) с двумя линейными квантователями [3, 4] и искусственный нейрон с двумя эллиптическими квантователями [5]. Так же был создан фрактально-корреляционный функционал, построенный на упорядочивании по возрастанию одной из, анализируемых переменных [6]. В итоге, на текущий момент, мы способны создать сеть из 17 нейронов, обобщающую 17 известных на сегодняшний день статистических критериев проверки гипотезы независимости.

Синтез четырех новых статистических критериев для проверки гипотезы независимости данных малых выборок

Так как увеличение числа статистически критериев (числа искусственных нейронов) выгодно, попытаемся синтезировать еще четыре новых статистических критерия. При этом будем синтезировать новые статистические критерии (новые нейроны), опираясь на классическую формулу Пирсона-Эджуорта-Эудлона (1), используемую при статистических оценках более 130 лет. Одна из модификаций формулы (1), может быть, построена путем отказа от нормирования данных через деление на стандартные отклонения (деления на интегральные характеристики $\sigma(x)$, $\sigma(y)$). Будем в место интегральных нормирующих статистик использовать частные статистики, вычисляемые для каждой точки. В качестве частных нормирующих статистик будем использовать квадраты расстояния от каждой точки выборки до центра двухмерного распределения данных [7]:

$$\tilde{r}(x,y)=\frac{1}{16}\sum_{i=1}^{16}\frac{\left(E(x)-x_i\right)\cdot\left(E(y)-y_i\right)}{\left(E(x)-x_i\right)^2+\left(E(y)-y_i\right)^2}. \qquad (3)$$

Результаты численного моделирования нового критерия (3) приведены на рис. 2.

Если сравнивать данные рис. 1 и рис. 2 становится очевидным снижение мощности нового критерия по сравнению с классическим критерием (1). Для классической формулы одинаковые вероятности ошибок первого и второго рода ее нейрона составляют значения $P_1 \approx P_2 \approx 0{,}146$. Для модифицированной формулы (2) равновероятные

ошибки ее нейрона выше $P_1 \approx P_2 \approx 0{,}211$. Мы наблюдаем снижение мощности нового статистического критерия по сравнению с классикой в 1,45 раза. Однако принципиальную важность имеет то, что два нейрона, эквивалентные вычислениям по двум формулам, имеют существенную независимую компоненту их выходных состояний $corr(r, \tilde{r}) = 0{,}75$. Если бы отклики двух нейронов были бы полностью зависимы, то использовать такие нейроны в одной нейросети было бы не целесообразно.

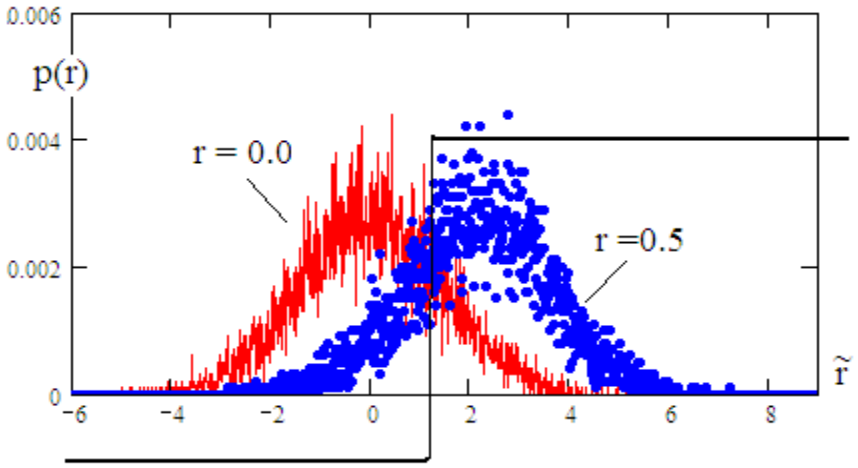

Рис. 2. Распределение оценок коэффициентов корреляции, вычисленных по формуле, являющейся аналогом классической формулы

Возможны и иные варианты модификации классической формулы (1). Проведенные исследования показали, что приемлемые результаты дают следующие модификации:

$$\tilde{r}_1(x,y) = \frac{1}{16}\sum_{i=1}^{16} \frac{(\min(x) - x_i)\cdot(\min(y) - y_i)}{(\min(x) - x_i)^2 + (\min(y) - y_i)^2}, \qquad (4)$$

$$\tilde{r}_2(x,y) = \frac{1}{16}\sum_{i=1}^{16} \frac{(\min(x) - x_i)\cdot(\max(y) - y_i)}{(\min(x) - x_i)^2 + (\max(y) - y_i)^2}, \qquad (5)$$

$$\tilde{r}_3(x,y) = \frac{1}{16}\sum_{i=1}^{16} \frac{(\max(x) - x_i)\cdot(\max(y) - y_i)}{(\max(x) - x_i)^2 + (\max(y) - y_i)^2}. \qquad (6)$$

Численный эксперимент по моделированию новых статистических критериев (4)–(6) дает распределения, похожие на распределения рис. 2. Равные вероятности ошибок первого и второго рода $P_1 \approx P_2$ и коэффициенты корреляционной сцепленности выходных состояний, пяти рассмотренных выше нейронов, приведены в табл. 1.

Таблица 1

Статистические параметры нейросетевого обобщения пяти критериев

Формула	$P_1 \approx P_2$			Матрица коэффициентов корреляции				
			Формула	r (1)	$\tilde{r}$ (2)	$\tilde{r}_1$ (3)	$\tilde{r}_2$ (4)	$\tilde{r}_3$ (5)
R (1)	0,146		r(1)	**1**	0,75	0,338	0,344	0,316
$\tilde{r}$ (2)	0,211		$\tilde{r}$ (2)	0,75	**1**	0,317	0,324	0,295
$\tilde{r}_1$ (3)	0,334		$\tilde{r}_1$ (3)	0,338	0,317	**1**	–0,17	0,089
$\tilde{r}_2$ (4)	0,395		$\tilde{r}_2$ (4)	0,344	0,324	–0,17	**1**	–0,192
$\tilde{r}_3$ (5)	0,328		$\tilde{r}_3$ (5)	0,316	0,295	0,089	–0,192	**1**

Таким образом, к 17 известным статистическим критерием мы имеет возможность добавить еще 4 статистических критериев, получив тем самым нейросеть обобщающую 21 известный на сегодня статистический критерий.

Иерархическая структура самокорректирующихся кодов, ориентированных применение в нейросетевых обобщениях

Следует отметить, что объединение в одну группу всех пяти, рассмотренных критериев не является оптимальным. Проведенные исследования показали, что в одну группу выгодно объединять статистические критерии с низкой взаимной коррелированностью и близкими вероятностями ошибок первого и второго рода [8]. То есть в одну общую группу корректировки кодов нам целесообразно объединять три критерия $\tilde{r}_1$, $\tilde{r}_2$, $\tilde{r}_3$ так как они имеют примерно одинаковые вероятности ошибок первого и второго рода {0,334, 0,395, 0,328} при среднем значении модулей коэффициентов корреляции $E(|r|) \approx$ $\approx 0{,}227$. Оптимизированная иерархическая структура организации данных самокорректирующегося кода для новых статистических критериев приведена на рис. 3.

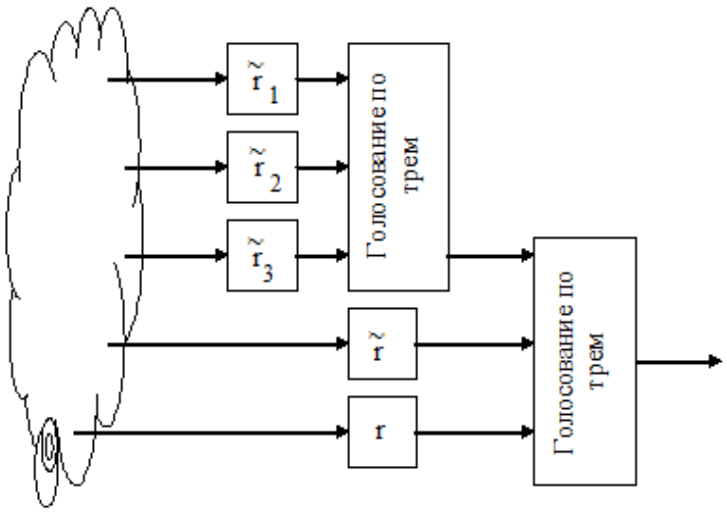

Рис. 3. Оптимизированная иерархическая структура организации самокорректирующегося кода для пяти рассматриваемых статистических критериев

Проведенное численное моделирование показало повышение корректирующей способности, рассмотренного самокорректирующегося иерархического кода на 21 % в сравнении с кодами без иерархии [8, 9].

Список литературы

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

2. Википедия. URL: https:// wikipedia.org/wiki/Корреляция

3. Волчихин В. И., Иванов А. И., Сериков А. В., Серикова Ю. И. Квантовая суперпозиция дискретного спектра состояний математической молекулы корреляции для малых выборок биометрических данных // Вестник Мордовского университета. 2017. Т. 27, № 2. С. 230–243. URL: http://vestnik.mrsu.ru/index.php/ru/articles2/51-17-2/320-10-15507-0236-2910-027-201702-07

4. Волчихин В. И., Иванов А. И., Сериков А. В., Серикова Ю. И. Тестирование аналогового и квантового оракулов линейной вычислительной сложности, предсказывающих значения коэффициента корреляции на малой выборке в 32 опыта // Известия высших учебных заведений. Поволжский регион. Технические науки. 2017. № 3. С. 70–80. URL: https://izvuz_tn.pnzgu.ru/tn6317

5. Сериков А. В., Качалин С. В. Корреляционная молекула с эллиптическими квантователями для вычислений на малых обучающих выборках // Безопасность информационных технологий : сб. науч. ст. по материалам I Всерос. науч.-техн. конф. (г. Пенза, 24 апреля 2019 г.). Пенза, 2019. С. 123–129. URL: https://www.elibrary.ru/item.asp?id=40845233&pf=1

6. Волчихин В. И., Ахметов Б. Б., Иванов А. И., Серикова Ю. И. Фрактально-корреляционный функционал, используемый при поиске пар слабо зависимых биометрических данных в малых выборках // Известия высших учебных заведений. Поволжский регион. Технические науки. 2016. № 4 (40). С. 27–36. URL: https://izvuz_tn.pnzgu.ru/tn3416

7. Volchikhin V. I., Ivanov A. I., Zolotareva1 T. A., Skudnev D. M. Synthesis of four new neuro-statistical tests for testing the hypothesis of independence of small samples of biometric data // APITECH III 2021. Journal of Physics: Conference Series. 2021. Vol. 2094. P. 032013. doi:10.1088/1742-6596/2094/3/032013

8. Иванов А. И., Золотарева Т. А. Искусственный интеллект в защищенном исполнении: синтез статистико-нейросетевых автоматов многокритериальной проверки гипотезы независимости малых выборок биометрических данных : препринт. Пенза : Изд-во ПГУ, 2020. 105 с.

9. Безяев А. В. Биометрико-нейросетевая аутентификация: обнаружение и исправление ошибок в длинных кодах без накладных расходов на избыточность : препринт. Пенза : Изд-во ПГУ, 2020. 40 с.

10. ГОСТ Р 52633.5–2011. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа.

Для цитирования: Золотарева Т. А., Безяев А. В., Олейник Ю. И. Иерархическая структура связей самокорректирующихся кодов, ориентированных на нейросетевое обобщение множества статистических критериев проверки гипотезы независимости малых выборок // Безопасность информационных технологий : сб. науч. ст. по материалам IV Всерос. науч.-техн. конф. : в 2 т. Пенза : Изд-во ПГУ, 2022. Т. 1. С. 18–26.

УДК 519.24; 57.017

ОЦЕНКА ПОТЕНЦИАЛЬНОГО РОСТА ЧИСЛА ВЫХОДНЫХ СОСТОЯНИЙ МНОГОУРОВНЕВЫХ КВАНТОВАТЕЛЕЙ ДЛЯ СЕТЕЙ КВАДРАТИЧНЫХ НЕЙРОНОВ ПРИ ИХ ПРОГРАММНОМ ВОСПРОИЗВЕДЕНИИ В МАССОВЫХ КОНТРОЛЛЕРАХ SIM-КАРТ

Ю. И. Серикова[1], Е. А. Малыгина[2], Т. А. Золотарева[3]

[1,2]Пензенский государственный университет, г. Пенза
[3]Липецкий государственный педагогический университет имени П. П. Семенова-Тян-Шанского, г. Липецк

Аннотация. Рассматривается перспектива перехода от сетей искусственных нейронов с накоплением данных в линейном пространстве и накоплением данных квадратичными нейронами. Показано, что переход от линейных к квадратичным нейронам позволяет снизить число входов с 16 до 4 при анализе динамики рукописного почерка. При этом обычными радиальными нейронами удается достичь показателя от 5- до 9-кратного виртуального распараллеливания вычислений за счет применения выходных квантователей. Переход к использованию нейронов Махаланобиса с обучением через выбор независимых входных данных позволяет повысить в среднем число выходных состояний квантователей до 13.

Ключевые слова: обогащение данных, искусственные нейроны, многоуровневые квантователи

EVALUATION OF THE POTENTIAL GROWTH IN THE NUMBER OF OUTPUT STATES OF MULTILEVEL QUANTIZERS FOR NETWORKS OF QUADRATIC NEURONS DURING THEIR SOFTWARE REPRODUCTION IN MASS SIM-CARD CONTROLLERS

Yu. I. Serikova[1], E. A. Malygina[2], T. A. Zolotareva[3]

[1,2]Penza State University, Penza
[3]Lipetsk State Pedagogical University named after P. P. Semenov-Tyan-Shansky, Lipetsk

Abstract. The perspective of transition from networks of artificial neurons with data accumulation in linear space and data accumulation by quadratic neurons is considered. It is shown that the transition from linear to quadratic neurons makes it

© Серикова Ю. И., Малыгина Е. А., Золотарева Т. А., 2022

possible to reduce the number of inputs from 16 to 4 when analyzing the dynamics of handwriting. At the same time, ordinary radial neurons manage to achieve an indicator from 5 to 9 times the virtual parallelization of calculations through the use of output quantizers. The transition to the use of Mahalanobis neurons with learning through the choice of independent input data allows increasing the average number of output states of quantizers to 13.

Keywords: data enrichment, artificial neurons, multilevel quantizers

При использовании сетей искусственных нейронов с линейным накоплением данных, например, реализованных в среде моделирования «БиоНейроАвтограф» [1], при автоматическом обучении по ГОСТ Р 52633.5–2011 необходимо использовать нейроны с 16 входами. Очевидно, что линейные нейроны (персептроны) плохо выделяют данные образ «Свой». На рис. 1 показано, что для выделения эллипса данных «Свой» требуется 4 персептрона в двухмерном пространстве. Для решения той же задачи достаточно одного квадратичного нейрона.

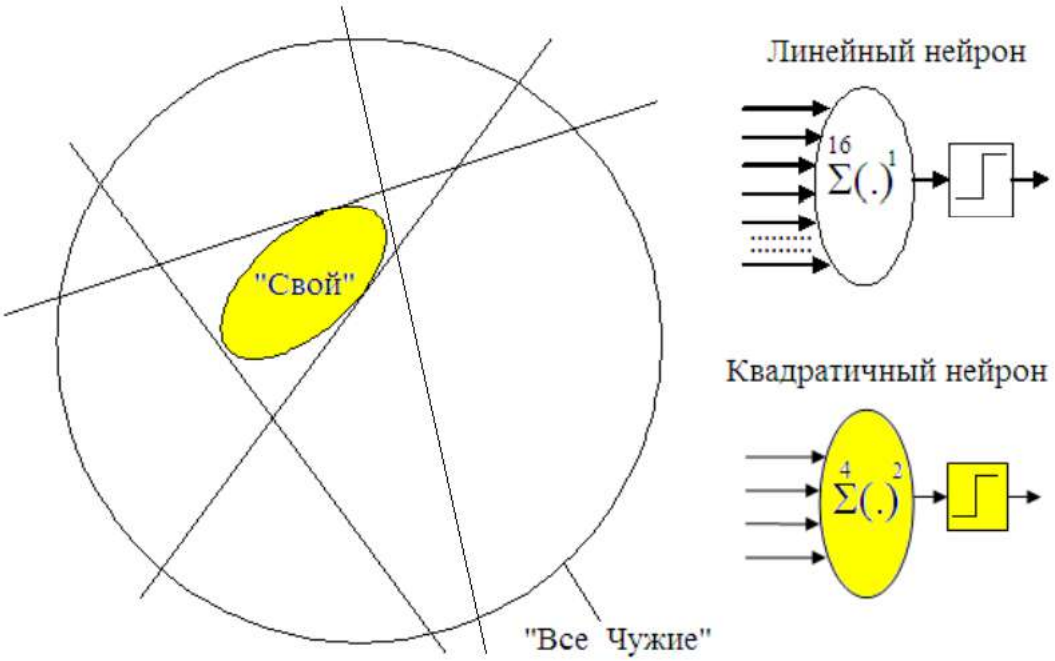

Рис. 1. Плохая выделимость образа «Свой» персептронами с линейными разделяющими функциями в сравнении с квадратичными нейронами

Переход от использования персептронов с 16 входами к применению квадратичных нейронов позволяет снизить от 8 до 4 число входов. Это существенно меньше по затратам памяти и доверенных вычислительных ресурсов при использовании массовых процессоров SIM-карт. В место того, чтобы воспроизводить 256 нейронов с 16 входами с бинарным квантователем мы можем перейти, например, к использованию 128 квадратичных нейронов с 7 или 8 входами

и троичными (четверичными) выходными квантователями. На рис. 2 представлена ситуация, перехода к 5-ти уровневым квантователям с четырьмя порогами $k_1 = 2{,}5$, $k_2 = 7$, $k_3 = 14$, $k_4 = 33$.

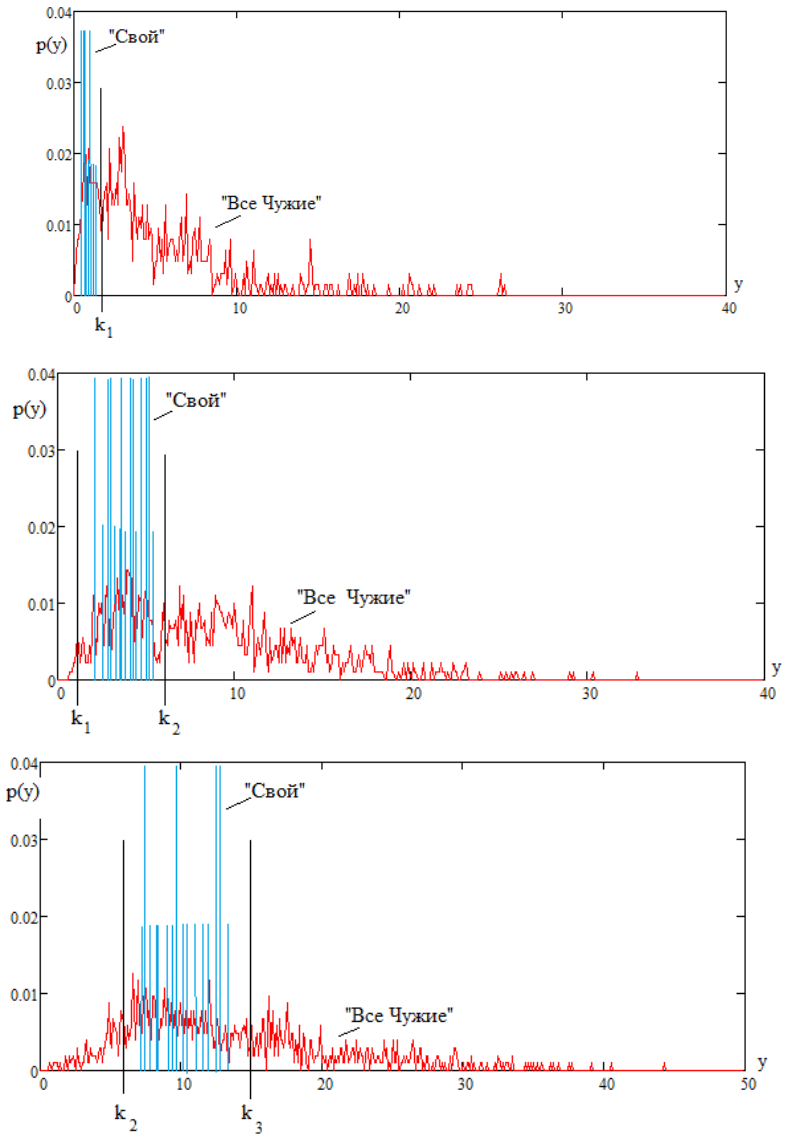

Рис. 2. Переход к использованию многоуровневых квантователей с четырьмя порогами сравнения

Следует отметить, что алгоритм обучения квадратичных нейронов с многоуровневым квантованием существенно усложняется [2, 3], однако он остается автоматическим (не требуется вмешательство человека).

Очевидным является так же то, что переход к многоуровневому квантованию выходных данных квадратичных нейронов оказывается технически возможным из-за соотношения вероятностей попадания в область «Свой», расположенную внутри существенно большего множества «Все Чужие». Чем меньше область «Свой» по объему

в сравнении с объемом области «Все Чужие», тем больше выходных состояний может иметь квантователь искусственного нейрона. Вычислить аналитически соотношение этих объемов затруднительно. При этом численно их оценить не сложно, достаточно оценить усредненную вероятность попадания хотя бы в первый в интервал квантования «Свой» данных множества образов «Все Чужие» для достаточно представительного числа нейронов.

Очевидным является так же то, что вероятностное соотношение многомерных объемов «Свой» и «Все Чужие» должны монотонно расти по мере увеличения числа входов у нейронов. Пользуясь этим, мы можем указать для той или иной биометрической технологии соотношение между числом входов у квадратичных нейронов и допустимым числом уровней его выходного квантователя.

В связи с выше сказанным, проведем численный эксперимент с использованием реальных данных среды «БиоНейроАвтограф» [1]. Результаты моделирования для квадратичных нейронов с 2, 3, 4, 5, 6, 7, 8 входами приведены в табл. 1.

Таблица 1

Оценка числа допустимого значения интервалов квантования для квадратичных нейронов с разным числом входов для рукописного образа «Пенза»

Число входов квадратичного нейрона	2	3	4	5	6	7	8
Вероятность P («Свой-1»)	0,45	0,37	0,31	0,228	0,162	0,96	0,78
Число квантов без округления	2,222	2,70	3,23	4,38	6,17	10,4	12,8
Округленное число квантов	«2»	«3»	«3»	«4»	«6»	«10»	«13»

Из табл. 1 видно, что число выходных состояний у квантователя быстро (квадратично) увеличивается, если расстояния между порогами выставлять линейно (пропорционально вероятность попадания данных «Все Чужие» в первый интервал образа «Свой»).

На самом деле интервалы между порогами квадратично растут по мере увеличения их номера, как это показано, на рис. 2. Это, видимо, делает рост числа уровня квантования линейным. То есть, увеличения числа входов у квадратичного нейрона должно в первом приближении давать пропорциональный рост числа выходных состояний у выходного квантователя.

Более тщательная проверка этой гипотезы может быть осуществлена после создания среды моделирования «БиоНейроАвтографа» с квадратичными нейронами.

Список литературы

1. Иванов А. И., Захаров О. С. Среда моделирования «БиоНейро Автограф». URL: http://пниэи.рф/activity/science/noc/bioneuroautograph.zip

2. Малыгина Е. А. Биометрико-нейросетевая аутентификация: перспективы применения сетей квадратичных нейронов с многоуровневым квантованием биометрических данных : препринт. Пенза : Изд-во ПГУ, 2020. 114 с.

3. Иванов А. И., Безяев А. В., Малыгина Е. А., Серикова Ю. И. Второй национальный стандарт России по быстрому автоматическому обучению больших искусственных нейронных сетей на малых выборках биометрических данных // Безопасность информационных технологий : сб. науч. ст. по материалам I Всерос. науч.-техн. конф. (г. Пенза, 24 апреля 2019 г.). Пенза, 2019. С. 174–177.

Для цитирования: Серикова Ю. И., Малыгина Е. А., Золотарева Т. А. Оценка потенциального роста числа выходных состояний многоуровневых квантователей для сетей квадратичных нейронов при их программном воспроизведении в массовых контроллерах sim-карт // Безопасность информационных технологий : сб. науч. ст. по материалам IV Всерос. науч.-техн. конф. : в 2 т. Пенза : Изд-во ПГУ, 2022. Т. 1. С. 27–31.

УДК 519.24; 53; 57.017

БЫСТРОЕ АВТОМАТИЧЕСКОЕ ОБУЧЕНИЕ СЕТИ ПЕРСЕПТРОНОВ С ШЕСТИУРОВНЕВЫМИ ВЫХОДНЫМИ КВАНТОВАТЕЛЯМИ ЧЕРЕЗ ПОДБОР СОЧЕТАНИЙ МАТЕМАТИЧЕСКИХ ОЖИДАНИЙ БИОМЕТРИЧЕСКИХ ПАРАМЕТРОВ ПРИ ИХ УПОРЯДОЧИВАНИИ И ДРОБЛЕНИИ НА ТРИ ИНТЕРВАЛА

К. Н. Савинов[1], С. Е. Вятчанин[2], В. А. Цимбал[3]

[1,2]Пензенский государственный университет, г. Пенза
[3]Филиал Военной академии Ракетных войск стратегического назначения имени Петра Великого, г. Серпухов Московской области

Аннотация. Рассматривается перспектива перехода от сетей искусственных нейронов с бинарными выходными квантователями к нейронам с 6-уровневыми выходными состояниями. Показано, что быстрое обучение нейронов удается выполнить выбором биометрических параметров с математическими ожиданиями, расположенными в трех интервалах. При этом одни и те же входные биометрические данные приходится анализировать двумя разными искусственными нейронами. Первый нейрон должен иметь троичный выходной квантователь и обучаться алгоритмом ГОСТ Р 52633.5. Второй искусственный нейрон предложено использовать с бинарным выходным квантователем, который должен анализировать размах входных данных.

Ключевые слова: автоматическое обучение искусственных нейронов, обогащение данных, многоуровневые квантователи

FAST AUTOMATIC LEARNING OF A NETWORK OF PERSEPTRONS WITH SIX LEVEL OUTPUT QUANTIZERS THROUGH THE SELECTION OF COMBINATIONS OF MATHEMATICAL EXPECTATIONS OF BIOMETRIC PARAMETERS WHEN THEIR ORDERING AND SPLITTING INTO THREE INTERVALS

K. N. Savinov[1], S. E. Vyatchanin[2], V. A. Tsymbal[3]

[1,2]Penza State University, Penza
[3]Branch of Military Academy of Strategic Rocket Troops after Peter the Great, Serpukhov, Moscow region

Abstract. The perspective of transition from networks of artificial neurons with binary output quantizers to neurons with 6-level output states is considered. It is shown

© Савинов К. Н., Вятчанин С. Е., Цимбал В. А., 2022

that fast learning of neurons can be performed by choosing biometric parameters with mathematical expectations located in three intervals. In this case, the same input biometric data has to be analyzed by two different artificial neurons. The first neuron must have a ternary output quantizer and be trained by the GOST R 52633.5 algorithm. The second artificial neuron is proposed to be used with a binary output quantizer, which should analyze the range of input data.

Keywords: automatic training of artificial neurons, data enrichment, multilevel quantizers

На сегодняшний день остро стоит вопрос об использовании приложений искусственного интеллекта в защищенном исполнении. В этом отношении единственным на сегодня легитимным решением является применение сети искусственных нейронов, обученных по ГОСТ Р 52633.5 [1]. При этом таблицы связей нейронов и таблицы весовых коэффициентов нейронов должны быть защищены криптографическими механизмами по соответствующей технической спецификации [2].

Опыт применения сетей бинарных персептронов, обученных по ГОСТ Р 52633.5 [1] показал, что они при шифровании их таблиц [2] существенно утрачивают качество принимаемых ими решений по соотношению вероятностей ошибок первого и второго рода. В связи с этим актуальной оказывается задача перехода к использованию более сложных искусственных нейронов с многоуровневым квантованием выходных данных [3].

Одной из причин подобных усложнений является низкая эффективность использования биометрических данных близких к данным среднего статистического образа «Чужой». На рис. 1 приведено реальное распределения биометрических параметров динамики воспроизведения парольного слова «*Пенза*» [4].

Из рис. 1 видно, что значения математических ожиданий имеют распределение, хорошо описываемое смесью нормального распределения и равномерного распределения. При этом все данные мы можем разделить на три группы. Вероятности попадания в каждую из этих групп одинаковые и составляют – 0,333. В левую группу попадают математические ожидания, расположенные в интервале от –20 до –2,4, соответствующие биометрическим данным с высоким уровнем отрицательной уникальности. В центральную группу попадают биометрические параметры с низкой уникальностью, характерные для динамики среднестатистического почерка. В правую группу попадают математические ожидания, расположенные в интервале от +2,5 до 20, соответствующие биометрическим данным с высоким уровнем положительной уникальности.

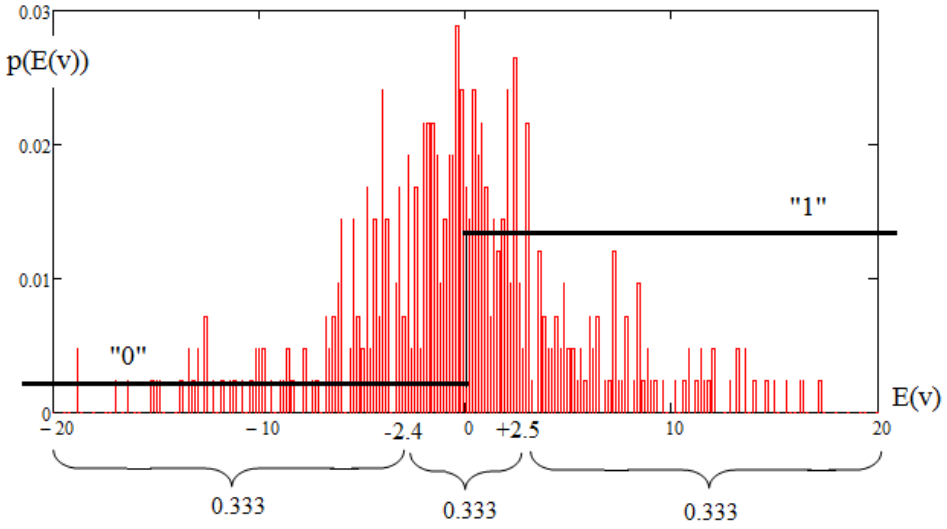

Рис. 1. Распределение значений 416 математических ожиданий биометрических параметров $E(v)$ динамики воспроизведения рукописного пароля «*Пенза*»

Следует так же подчеркнуть, что правая и левая группы биометрических параметров оказывают существенное влияние на состояние бинарных персептронов, обученных алгоритмом по ГОСТ Р 52633.5 [1]. Это связано с вычислением модулей весовых коэффициентов, совпадающими с качеством данных и вычисленным по формуле:

$$|\mu_i| = \frac{|E(v_i)|}{\sigma(v_i)} = q_i, \qquad (1)$$

где $|E(v_i)|$ – модуль математического ожидания; $\sigma(v_i)$ – стандартное отклонение i-го биометрического параметра.

Для правой и левой группы с большими значениями модулей математических ожиданий биометрических параметров модули весовых коэффициентов нейронов будут значимыми, попадая в интервал от 1 до 100. Напротив, биометрические параметры с математическими ожиданиями, попадающими в центральный узкий интервал, дают малозначимые модули весовых коэффициентов, попадающие в интервал от 0 до 1. В связи с этим возникает задача сделать биометрические параметры близкие к средним статистическим значимыми.

Алгоритм действующего стандарта по автоматическому обучению бинарных персептронов ГОСТ Р 52633.5 [1] предполагает случайный выбор биометрических параметров. В рассматриваемом случае, анализа данных для рукописного пароля, воспроизводимого ручкой на графическом планшете, требуются бинарные персептроны

с 16 входами, т.е. для каждого нейрона должна выполняться псевдослучайная выборка по 16 из 416 параметров с номерами $n = \{1, 2, \ldots, 416\}$.

В нашем случае используется похожий алгоритм псевдослучайного выбора входных связей персептронов. Он построен на предварительном упорядочивании биометрических параметров по их математическому ожиданию с номерами m. Однако итоговые входные связи нейронов кажутся случайными для стороннего наблюдателя, не знающего качество данных биометрического образа «Свой». Пример соотношения исходных номеров входов нейросети и тех же номеров после упорядочивания данных подтверждается табл. 1.

Таблица 1

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	::::
m	415	3	232	258	224	171	35	408	135	354	307	134	143	31	95	::::
	::::: :::::	::::: :::::	::::::: :::::::	::::::: :::::::	::::::: :::::::	::::::: :::::::	::::: :::::	::::::: :::::::	::::::: :::::::	::::::: :::::::	::::::: :::::::	::::::: :::::::	::::::: :::::::	::::::: :::::::	::::::: :::::::	::::: :::::
n	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415
m	350	363	198	251	237	84	20	99	226	257	136	347	402	260	67	375

То, что упорядочивание данных по математическому ожиданию слабо связано с порядком данных, поступающих на вход всей нейросети легко проверяется вычислением корреляции $\mathrm{corr}(n, m) = -0.013$.

Пример реализации работы алгоритмов сортировки при обучении троичного персептрона приведен на рис. 2.

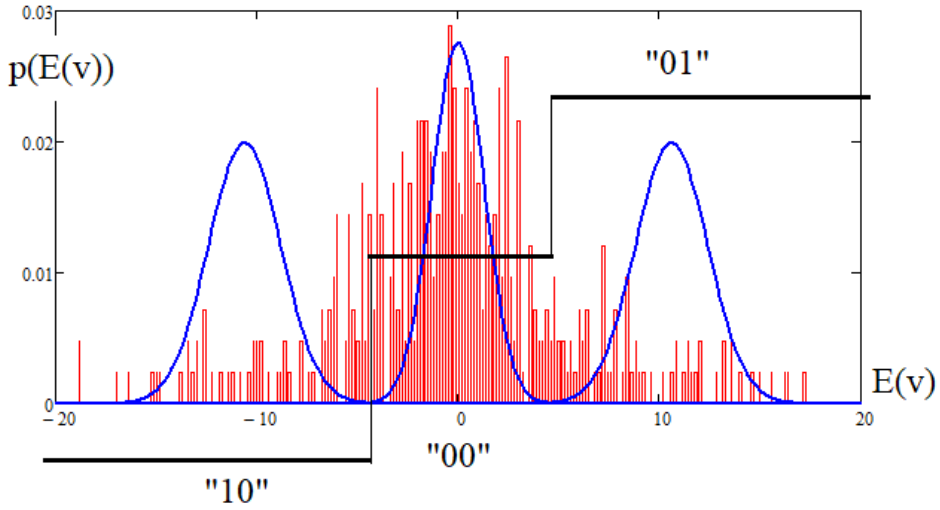

Рис. 2. Распределения данных, на входе и выходе троичного квантователя, обученного персептрона

Наиболее простым для понимания является обучение сумматора персептрона, ориентированного попадание данных в центральный участок квантователя. Для этого достаточно, как и в предшествующем стандарте, случайно выбрать 8 параметров из возможных математических ожиданий, лежащих в интервале от –2,4 до +2,5 с номерами после упорядочивания m = {139, 140, …, 287}.

По этим данным нужно для каждого из 8-ми параметров с уже заданными номерами необходимо найти соседние параметры справа или слева (m±1). Выбор знака (право или лево) выполняется случайно.

В конечном итоге мы получаем обученный «центральный» нейрон с весовыми коэффициентами, вычисленными по формуле (1). При этом число отрицательных знаков и положительных знаков для этих весовых коэффициентов будут случайными

Очевидным является то, что с ростом числа входов у нейронов происходит рост нормализации данных на выходах их сумматоров. Так как в «центре» распределения данных, они наиболее плотно упакованы, то «центральное» нормальное распределение дает выходное распределение сумматоров с наименьшим стандартным отклонением (рис. 2).

Обучение (настройка) персептрона левого интервала (рис. 1) ведется случайным выбором 8-ми биопараметров с номерами m = = {0, 1, …, 138}. Далее параметры с этими номерами подаются на сумматор левого персептрона с положительными знаками. С отрицателными знаками на этот же сумматор подают 8, случайно выбранных параметров с номерами из правой последовательности m = {287, 288, …, 415}. Значение модуля весового коэффициента выбирается вычислением по формуле (1).

Если необходимо при обучении персептрона попадать в правый интервал (рис. 2), откликаясь на данные «Свой», то 8 случайно, выбранных параметров, из последовательности m = {0, 1, …, 138} следует суммировать с отрицательными весовыми коэффициентами. Напротив 8 других случайно, выбранных параметра из последовательности m = {287, 288, …, 415} должны суммироваться «правым» персептроном с положительными знаками.

Заметим, что номера суммируемых данных выбираются случайно и как следствие распределение данных на выходе сумматора нормализуется, но положением трех математических ожиданий и трех стандартных отклонений, могут иметь некоторую неопределенность. Если такая неопределенность оказывается негативной, то ее устранение ведется увеличением входных данных сумматора.

Допустимо при обучении персептрона увеличивать число входных биометрических параметров. Рост числа биометрических параметров всегда приводит к стабилизации математического ожидания и снижению стандартного отклонения данных на выходе сумматора.

Заметим, что, описанный выше алгоритм обучения может иметь для образа «Свой» малый размах анализируемых данных или большой размах анализируемых нейроном данных. В связи с этим мы имеем возможность параллельно применить второй искусственный нейрон, анализирующий размах входной выборки биометрических данных [5, 6]:

$$\begin{cases} x \leftarrow sort(x) \\ d \leftarrow \dfrac{x_{16}+x_{15}}{2} - \dfrac{x_1+x_2}{2} \\ z(d) \leftarrow "0"\ if\ \ d \le k \\ z(d) \leftarrow "1"\ if\ \ d > k \\ P_1 \approx P_2 \approx P_{EE} \approx 0.5 \end{cases}. \qquad (2)$$

Размах выборки – d зависит от того, как формировалась эта выборка. Если эта выборка формировалась только из данных одного интервала рис. 1 и математические ожидания биометрических параметров близки, размах выборки мал. Ситуация меняется, если данные выбираются из правого и левого интервала (центральный интервал рисунка 1 не используется). В этом случае размах данные выборки оказывается значительным. Это в конечном случае позволяет получить шесть выходных состояний двух квантователей двух рассматриваемых нейронов. Выходные состояния троичного и бинарного квантователей должны свертываться некоторым логическим автоматом.

Список литературы

1. ГОСТ Р 52633.5–2011. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа.

2. Техническая спецификация «Системы обработки информации. Криптографическая защита информации. Защита нейросетевых биометрических контейнеров с использованием криптографических алгоритмов» : принята 19.11.2020 на XXV заседании технического комитета № 26.

3. Малыгина Е. А. Биометрико-нейросетевая аутентификация: перспективы применения сетей квадратичных нейронов с многоуровневым квантованием биометрических данных : препринт. Пенза : Изд-во ПГУ, 2020. 114 с.

4. Иванов А. И., Захаров О. С. Среда моделирования «БиоНейроАвтограф». URL: http://пниэи.рф/activity/science/noc/ bioneuroautograph.zip

5. Иванов А. И., Банных А. Г., Куприянов Е. Н. [и др.]. Коллекция искусственных нейронов эквивалентных статистическим критериям для их совместного применения при проверке гипотезы нормальности малых выборок биометрических данных // Безопасность информационных технологий : сб. науч. ст. по материалам I Всерос. науч.-техн. конф. (г. Пенза, 4 апреля 2019 г.). Пенза, 2019. С. 156–164.

6. Иванов А. И. Искусственные математические молекулы: повышение точности статистических оценок на малых выборках (программы на языке MathCAD) : препринт. Пенза : Изд-во ПГУ, 2020. 36 с.

Для цитирования: Савинов К. Н., Вятчанин С. Е., Цимбал В. А. Быстрое автоматическое обучение сети персептронов с шестиуровневыми выходными квантователями через подбор сочетаний математических ожиданий биометрических параметров при их упорядочивании и дроблении на три интервала // Безопасность информационных технологий : сб. науч. ст. по материалам IV Всерос. науч.-техн. конф. : в 2 т. Пенза : Изд-во ПГУ, 2022. Т. 1. С. 32–38.

УДК 57.017

ОПТИМИЗАЦИЯ ПАРАМЕТРОВ ПОЛИНОМИНАЛЬНОГО КРИТЕРИЯ ЛЕЖАНДРА, ОРИЕНТИРОВАННОГО НА ПРОВЕРКУ ГИПОТЕЗЫ РАВНОМЕРНОГО РАСПРЕДЕЛЕНИЯ ДАННЫХ МАЛЫХ ВЫБОРОК

Е. Н. Куприянов

Пензенский государственный университет, г. Пенза

Аннотация. Показано, что ортогональные полиномы Лежандра второго порядка могут быть использованы при синтезе искусственных нейронов, ориентированных на проверку гипотезы равномерного распределения малых выборок биометрических данных. Положительный эффект возникает в случае нормирования, приводящего проверяемые биометрические данные к интервалу ортогональности классических полиномов Лежандра. Однако малые выборки не дают идеальных результатов. В связи с этим выполняется оптимизация параметров полинома второго порядка Лежандра.

Ключевые слова: оптимизация вычислений, обогащение данных, ортогональные полиномы Лежандра, искусственные нейроны

OPTIMIZATION OF THE PARAMETERS OF THE POLYNOMINAL LEGENDRE CRITERION ORIENTED TO TEST THE HYPOTHESIS OF THE UNIFORM DISTRIBUTION OF DATA IN SMALL SAMPLES

E. N. Kupriyanov

Penza State University, Penza

Abstract. It is shown that the Legendre orthogonal polynomials of the second order can be used in the synthesis of artificial neurons aimed at testing the hypothesis of a uniform distribution of small samples of biometric data. A positive effect arises in the case of normalization, which brings the verified biometric data to the orthogonality interval of classical Legendre polynomials. However, small samples do not give ideal results. In this regard, optimization of the parameters of the Legendre second-order polynomial is performed.

Keywords: computational optimization, data enrichment, Legendre orthogonal polynomials, artificial neurons

При развитии нейросетевой биометрии важным оказывается наличие эффективных процедур проверки гипотезы нормальности

© Куприянов Е. Н., 2022

и равномерности малых выборок. Стандартные статистические критерии [1, 2] проверки гипотез нормальности и равномерности распределения данных хорошо работают только при больших выборках в 200 и более опытов. Можно пойти по пути нейросетевого объединения нескольких статистических критериев [3–5], однако можно пойти и по иному ортогонализации при синтезе новых статистических критериев [6, 7].

В частности, при синтезе новых статистических критериев могут быть использованы полиномы Лежандра второго порядки, ортогональные на интервале от –1 до +1. На рис. 1 представлена программная реализация, воспроизводящая статистический критерий модуля полинома Лежандра второго порядка.

$$
\begin{aligned}
&n := 16 \qquad a := 1\\
&xxx(rr) := \left|\begin{array}{l}
xn \leftarrow \mathrm{sort}(\mathrm{rnorm}(n, 0, 1 + rr))\\
xr \leftarrow \mathrm{sort}(\mathrm{runif}(n, -3, 3))\\
xn \leftarrow \left(xn - xn_0\right)\\
xr \leftarrow \left(xr - xr_0\right)\\
xn \leftarrow \dfrac{2 \cdot xn}{xn_{15}} - 1\\
xr \leftarrow \dfrac{2 \cdot xr}{xr_{15}} - 1\\
xn \leftarrow \left|\displaystyle\sum_{i=0}^{n-1} \left[\left[3 \cdot \left[(xn)_i\right]^2 - a\right]^2 \cdot \frac{1}{16}\right]\right|\\
xr \leftarrow \left|\displaystyle\sum_{i=0}^{n-1} \left[\left[3 \cdot \left[(xr)_i\right]^2 - a\right]^2 \cdot \frac{1}{16}\right]\right|\\
\begin{pmatrix} xn \\ xr \end{pmatrix}
\end{array}\right.
\end{aligned}
$$

Рис. 1. Программное обеспечение, воспроизводящее статистический критерий модуля полинома Лежандра второго порядка

Результаты численного моделирования нового типа нейронов представлены на рис. 2.

Из рис. 2 видно, что даже для малой выборки из 16 опытов удается хорошо разделить отклики на распределенные нормально данные и на распределенные равномерно данные. Если принять порог сравнения $L(x) = 2$, то равновероятные ошибки первого и второго рода $P_1 = P_2 = P_{EE} = 0{,}366$. Эта оценка получена в рамках предположения полного совпадения, исследуемого критерия с классическим

полиномом Лежандра второго порядка. Очевидно, что с ростом объема выборки вероятности ошибок будут падать. В связи с этим целесообразно оптимизировать параметры полинома Лежандра. Например, мы можем оптимизировать параметр «a».

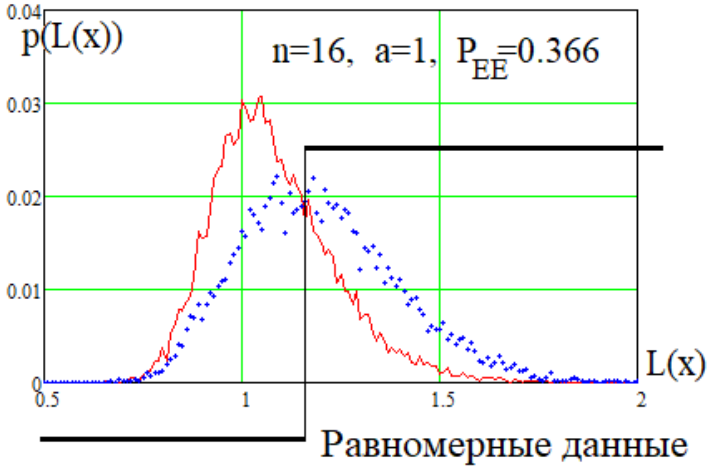

Рис. 2. Плотности распределения откликов статистического критерия классического полинома Лежандра второго порядка на малые выборки нормальных данных и равномерно распределенных данных

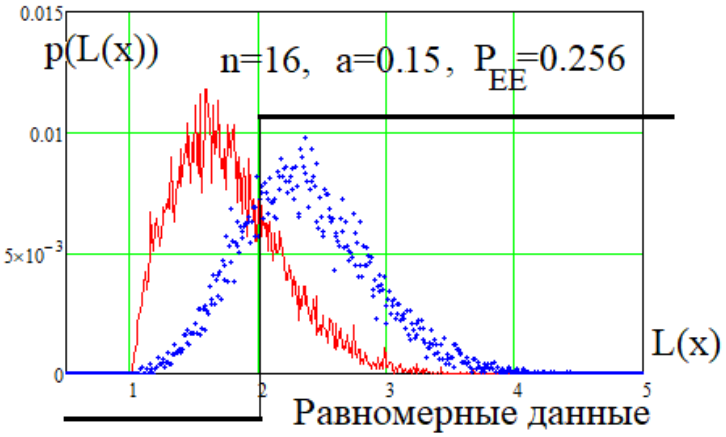

Рис. 3. Плотности распределения откликов статистического критерия не классического полинома Лежандра (существенно снижен постоянный член полинома)

Таким образом, для полиномов Лежандра второго порядка оптимизация постоянного члена позволяет снизить вероятность ошибок первого и второго рода на 41 % для выборок в 16 опытов. Видимо, и для статистических критериев, построенных на использовании полиномов Лежандра более высоких порядков, так же потребуется оптимизация их параметров под каждую выборку.

Список литературы

1. Р 50.1.037–2002. Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Ч. I. Критерии типа χ2. М. : Госстандарт России, 2001. 140 с.

2. Р 50.1.037–2002. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Ч. II. Непараметрические критерии. М. : Госстандарт России, 2002. 123 с.

3. Иванов А. И., Банных А. Г., Куприянов Е. Н. [и др.]. Коллекция искусственных нейронов эквивалентных статистическим критериям для их совместного применения при проверке гипотезы нормальности малых выборок биометрических данных // Безопасность информационных технологий : сб. науч. ст. по материалам I Всерос. науч.-техн. конф. (г. Пенза, 24 апреля 2019 г.). Пенза 2019. С. 156–164.

4. Волчихин В. И., Иванов А. И., Безяев А. В., Куприянов Е. Н. Нейросетевой анализ малых выборок биометрических данных с использованием хи-квадрат критерия и критериев Андерсона – Дарлинга // Инженерные технологии и системы. 2019. Т. 29, № 2. С. 205–217. URL: doi:https://doi.org/10.15507/2658-4123.029/2019.02.205-217

5. Иванов А. И. Искусственные математические молекулы: повышение точности статистических оценок на малых выборках (программы на языке MathCAD). Пенза : Изд-во ПГУ, 2020. 36 с.

6. Куприянов Е. Н., Иванов А. И. Ортогонализация статистико-нейросетевого анализа малых выборок биометрических данных на примере использования нейронов Лежандра в первом слое двухслойной сети искусственных нейронов // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. Пенза : Изд-во ПГУ, 2020. С. 67–72

7. Иванов А. И., Куприянов Е. Н. Защита искусственного интеллекта : ортогонализация статистико-нейросетевого анализа малых выборок биометрических данных : препринт. Пенза : Изд-во ПГУ, 2020. 72 с.

Для цитирования: Куприянов Е. Н. Оптимизация параметров полиномиального критерия Лежандра, ориентированного на проверку гипотезы равномерного распределения данных малых выборок // Безопасность информационных технологий : сб. науч. ст. по материалам IV Всерос. науч.-техн. конф. : в 2 т. Пенза : Изд-во ПГУ, 2022. Т. 1. С. 39–42.

УДК 519.23

ОПТИМИЗАЦИЯ ПРОЦЕДУРЫ СМЕЩЕНИЯ ВХОДНЫХ ДАННЫХ ДЛЯ НЕЙРОНОВ СРЕДНЕГО ГАРМОНИЧЕСКОГО, ИСПОЛЬЗУЕМЫХ ПРИ ПРОВЕРКЕ ГИПОТЕЗЫ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ МАЛЫХ ВЫБОРОК

В. С. Лукин[1], О. С. Лаута[2]

[1]Пензенский государственный университет, г. Пенза
[2]Военная академия связи имени Маршала Советского Союза С. М. Буденного Министерства обороны Российской Федерации, г. Санкт-Петербург

Аннотация. При автоматическом обучении нейросетевых преобразователей биометрии в код биометрической аутентификации по ГОСТ Р 52633.5 необходимо проверять гипотезу нормального распределения входных биометрических данных. Использование для этой цели хи-квадрат критерия дает значительный уровень ошибок при обучающих выборках в 16 примеров. При переходе к использованию критерия среднего гармонического удается снизить вероятность появления ошибок примерно в 3,7 раза. Оптимизация процедуры смещения входных данных позволяет дополнительно снизить вероятность появления ошибок, примерно в два раза.

Ключевые слова: многокритериальный нейросетевой статистический анализ, малые выборки, хи-квадрат критерий, критерий среднего гармонического, оптимизация параметров нормирования и смещения исходных данных

OPTIMIZATION OF THE PROCEDURE OF INPUT DATA BIAS FOR HARMONIC MEAN NEURONS USED IN VERIFICATION OF THE HYPOTHESIS OF THE NORMAL DISTRIBUTION OF SMALL SAMPLES

V. S. Lukin[1], O. S. Lauta[2]

[1]Penza State University, Penza
[2]Military Academy of Communications named after Marshal of the Soviet Union S. M. Budyonny of the Ministry of Defense of the Russian Federation, St. Petersburg

Abstract. When automatically training neural network converters of biometrics into a biometric authentication code according to GOST R 52633.5, it is necessary to check the hypothesis of the normal distribution of input biometric data. The use of the chi-squared test for this purpose gives a significant level of error with training samples of 16 examples. When switching to the use of the harmonic mean criterion, it is possible to reduce the probability of errors by about 3,7 times. Optimization of the input data shift procedure allows to additionally reduce the probability of errors, approximately by a factor of two.

© Лукин В. С., Лаута О. С., 2022

Keywords: multicriteria neural network statistical analysis, small samples, chi-square test, harmonic mean test, optimization of parameters of normalization and bias of initial data

Введение

Обучение нейронных сетей преобразованию биометрии в код аутентификации по ГОСТ Р 52633.5 [1] выполняется на 16 примерах образа «Свой». При этом «хорошие» биометрические данные имеют нормальное распределение, а «плохие» данные с грубыми ошибками имею распределение близкое к равномерному. В итоге при оценке качества малых обучающих выборок нужно проверять гипотезу нормального распределения малой выборки в 16 примеров.

Одним из очевидных способов проверки гипотезы нормальности является использование хи-квадрат критерия Пирсона. К сожалению, для малых выборок этот классический статистический критерий плохо работает. Эта ситуация иллюстрируется рис. 1.

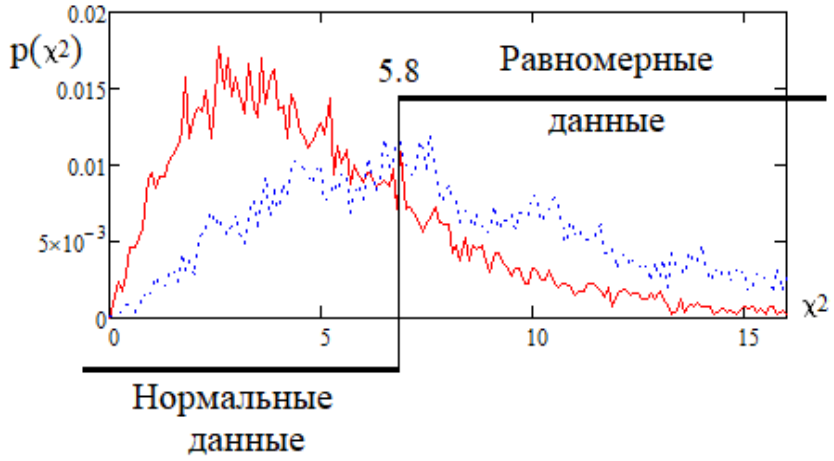

Рис. 1. Пример плохой линейной разделимости искусственным нейроном выходных состояний классического хи-квадрат критерия для малых выборок в 16 опытов

Очевидным является то, что для малых выборок вероятности ошибок первого и второго рода велики $P_1 = P_2 = P_{EE} \approx 0{,}330$. В связи с этим по стандартным рекомендациям [2] для приемлемых значений доверительных вероятностей критерий хи-квадрат должен применяться для выборок в 200 и более опытов. Это условие невыполнимо для нейросетевой биометрии.

Постановка задачи многокритериального нейросетевого анализа

Выходом из создавшегося положения является параллельное использование множества статистических критериев проверки гипотезы нормального распределения малых выборок. Так в прошлом

веке [3] было создано порядка 21 статистического критерия для проверки гипотезы нормальности. В работе [4] показано, что для каждого из известных статистических критериев проверки гипотезы нормальности может быть построен эквивалентный искусственный нейрон. Для этого достаточно на выходе свертывающего данные статистического функционала разместить бинарный квантователь, срабатывающий в точке равновероятных ошибок первого и второго рода. Для хи-квадрат нейрона квантователь должен срабатывать в точке $\chi^2 = 5{,}8$ как это показано на рис. 1.

Если мы каждый из 21 критерия [3] заместим своим эквивалентным искусственным нейроном, то на выходе нейросети получим выходной код с 21-кратной избыточностью. Устранение столь высокой избыточности может быть выполнено подсчетом состояний «0» и состояний «1». Итоговое решение должно приниматься по большинству выявленных в коде состояний.

Одним из ограничений такого мультикритериального анализа малых выборок является высокая корреляционная связанность классических статистических критериев. Учет влияния корреляционных связей [5] показывает, что для достижения доверительной вероятности 0,9 итогового нейросетевого решения необходимо использовать порядка 60 статистических критериев. Последнее означает, что в ближайшее время необходимо дополнительно синтезировать и оптимизировать порядка 40 новых статистических критериев.

Оптимизация смещения входных данных для критериев среднего гармонического

Одним из путей решения этой задачи является переход к использованию семейства статистических критериев среднего гармонического и среднего геометрического [6–9]. В частности, может быть использован искусственный нейрон среднего гармонического, полученный нормированием входных данных стандартным отклонением и их смещением в интервал от 1 до 9:

$$\begin{cases} x \leftarrow sort(x) \\ x \leftarrow \dfrac{x}{\sigma(x)} \\ x \leftarrow x - x_0 + 1{,}0 \\ \text{sga} \leftarrow \dfrac{\sqrt[16]{\prod_{i=0}^{15} x_i}}{E(x)} \\ z(sga) \leftarrow \text{"0"} \quad \text{if} \quad y > 0{,}908 \\ z(sga) \leftarrow \text{"1"} \quad \text{if} \quad y \le 0{,}908 \end{cases} . \qquad (1)$$

Численное моделирование соотношения (1) дает распределения выходных состояний искусственного нейрона отображенное на рис. 2.

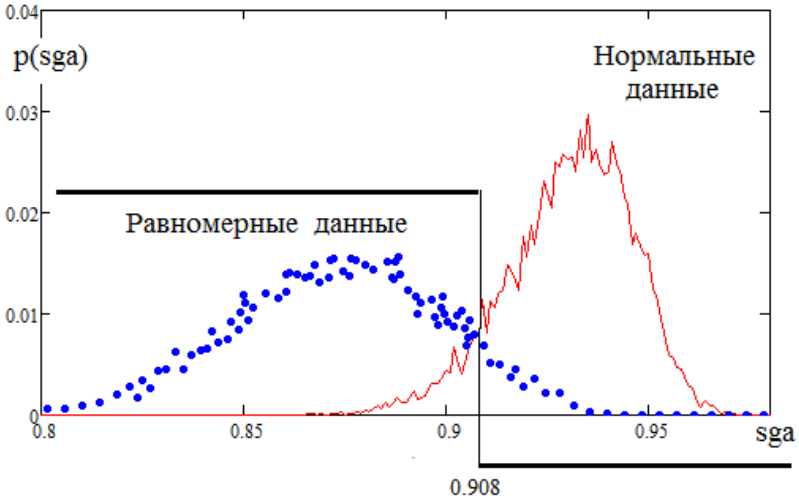

Рис. 2. Разделение нормальных и равномерных данных нейроном среднего гармонического для малых выборок в 16 опытов

Очевидным является то, что нормировка входных данных в (1) существенно влияет на результат вычислений. В связи с этим была предпринята попытка оптимизации ее параметров. Проведенные работы показали, что увеличение смещения в формуле (1) с 1,0 до 3,0 позволяет снизить вероятность появления ошибок первого и второго рода примерно в 2 раза с величины $P_1 = P_2 = P_{EE} \approx 0{,}088$ до величины $P_1 = P_2 = P_{EE} \approx 0{,}046$.

Соответствующий численный эксперимент выполнялся по программе на языке MathCAD, приведенной на рис. 3.

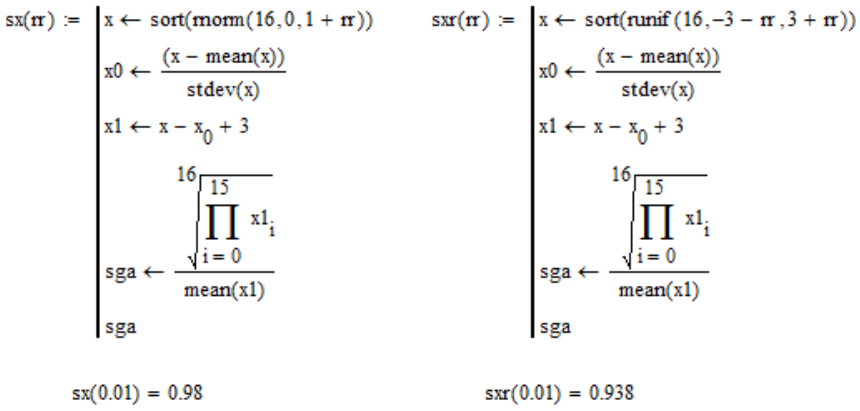

Рис. 3. Программа, воспроизводящая смещение данных в право на 3,0

Результатом смещения с 1,0 в формуле (1) до 3,0 является существенное изменение в право, наблюдаемых распределений откликов нового критерия, отображенных на рис. 4.

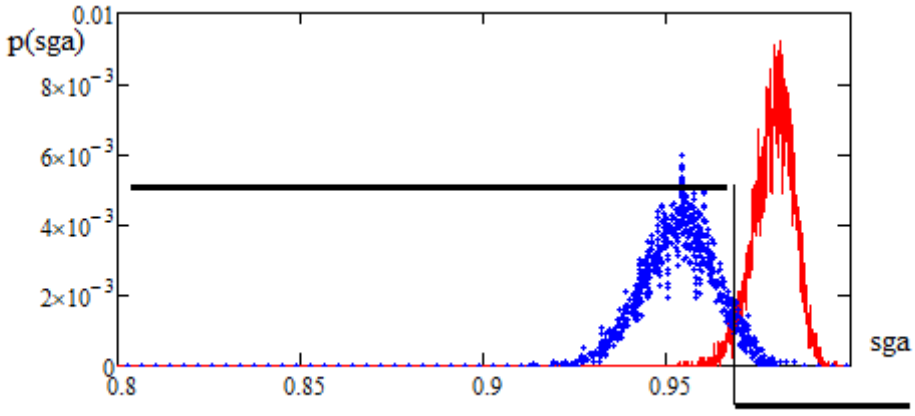

Рис. 4. Результаты численного моделирования при смещении 3,0

Дальнейшее увеличение показателя смещения с 3,0 до 15,0 показывает снижение значений равновероятных ошибок до величины $P_1 = P_2 = P_{EE} \approx 0{,}038$ при показателе смещения 8,7. Кривая изменения значений равновероятных ошибок первого и второго рода приведена на рис. 5.

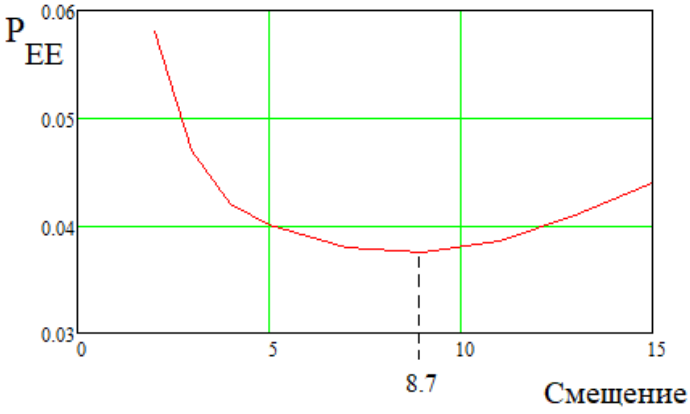

Рис. 5. Кривая изменения значений равновероятных ошибок первого и второго рода в зависимости от изменения смещения

Таким образом, оптимизация показателя смещения критерия среднего гармонического позволяет снизить вероятности ошибок в 2,3 раза. Это показывает то, на сколько важны исследования, связанные с оптимизацией вычислений.

Список литературы

1. ГОСТ Р 52633.5–2011. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа.

2. Р 50.1.037–2002. Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Ч. I. Критерии типа χ2. М. : Госстандарт России, 2001. 140 с.

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

4. Иванов А. И., Банных А. Г., Безяев А. В Искусственные молекулы, собранные из искусственных нейронов, воспроизводящих работу классических статистических критериев // Вестник Пермского университета. Сер.: Математика. Механика. Информатика. 2020. № 1 (48). С. 26–32.

5. Иванов А. И., Банных А. Г., Серикова Ю. И. Учет влияния корреляционных связей через их усреднение по модулю при нейросетевом обобщении статистических критериев для малых выборок // Надежность. 2020. № 20 (2). С. 28–34. URL: https://doi.org/10.21683/1729-2646-2020-20-2-28-34

6. Иванов А. И., Банных А. Г., Куприянов Е. Н. [и др.]. Коллекция искусственных нейронов эквивалентных статистическим критериям для их совместного применения при проверке гипотезы нормальности малых выборок биометрических данных // Безопасность информационных технологий : сб. науч. ст. по материалам I Всерос. науч.-техн. конф. (г. Пенза, 24 апреля 2019 г.). Пенза, 2019. С. 156–164.

7. Иванов А. И., Перфилов К. А., Лукин В. С. Нейросетевое обобщение семейства статистических критериев среднего геометрического и среднего гармонического для прецизионного анализа малых выборок биометрических данных // Информационно-управляющие телекоммуникационные системы, средства поражения и их техническое обеспечение : сб. науч. ст. Всерос. науч.-техн. конф. / под общ. ред. В. С. Безяева. Пенза : АО «НПП "Рубин"», 2019. С. 50–63.

8. Иванов А. И., Банных А. Г., Безяев А. В. Искусственный нейрон для контроля по критерию вариаций коэффициентов эксцесса малых выборок биометрических данных с нормальным распределением // Информационно-управляющие телекоммуникационные системы, средства поражения и их техническое обеспечение : сб. науч. ст. Всерос. науч.-техн. конф. / под общ. ред. В. С. Безяева. Пенза : АО «НПП "Рубин"», 2019. С. 84–94.

9. Лукин В. С. Сравнение мощности обычной и логарифмической форм статистических критериев среднего гармонического при использовании для проверки гипотезы нормального распределения данных малой выборки // Известия высших учебных заведений. Поволжский регион. Технические науки. 2020. № 4. С. 19–26.

Для цитирования: Лукин В. С., Лаута О. С. Оптимизация процедуры смещения входных данных для нейронов среднего гармонического, используемых при проверке гипотезы нормального распределения малых выборок // Безопасность информационных технологий : сб. науч. ст. по материалам IV Всерос. науч.-техн. конф. : в 2 т. Пенза : Изд-во ПГУ, 2022. Т. 1. С. 43–48.

УДК 519.24

МУЛЬТИКАТИВНОЕ ОБЪЕДИНЕНИЕ ДВУХ НОВЫХ СТАТИСТИЧЕСКИХ КРИТЕРИЕВ ДЛЯ ИХ ВЗАИМНОГО УСИЛЕНИЯ ПРИ ПРОВЕРКЕ ГИПОТЕЗЫ НОРМАЛЬНОСТИ ДАННЫХ МАЛОЙ ВЫБОРКИ

К. А. Перфилов[1], С. А. Полковникова[2], А. Ю. Малыгин[3], Е. Н. Куприянов[4]

[1,2,3,4] *Пензенский государственный университет, г. Пенза*

Аннотация. Рассматривается проблема статистического анализа малых выборок путем синтеза новых статистических критериев. Предложено перед расчетами выполнить операцию перемножения состояний двух относительно новых статистических критериев. Вероятность появления ошибок первого и второго рода классического хи-квадрат критерия при малой выборке в 16 опытов составляет 0,33, что недопустимо для практики. Новый статистический критерий при тех же условиях снижает вероятность ошибок до 0,075, что уже вполне допустимо для ряда приложений нейросетевой биометрии. При перемножении двух новых критериев удается снизить вероятности ошибок до вероятности 0,045.

Ключевые слова: статистический анализ малых выборок, проверка гипотезы нормальности, критерии среднего геометрического, дифференцирование случайных данных малых выборок

MULTICATED COMBINATION OF TWO NEW STATISTICAL CRITERIA FOR THEIR MUTUAL STRENGTHENING IN VERIFICATION OF THE HYPOTHESIS OF NORMALITY OF DATA OF A SMALL SAMPLE

K. A. Perfilov[1], S. A. Polkovnikova[2], A. Ya. Malygin[3], E. N. Kupriyanov[4]

[1,2,3,4] *Penza State University, Penza*

Abstract. The paper considers the problem of statistical analysis of small samples by synthesizing new statistical criteria. It is proposed to perform the operation of multiplying the states of two relatively new statistical criteria before the calculations. The probability of occurrence of errors of the first and second kind of the classical chi-square test with a small sample of 16 experiments is 0,33, which is unacceptable for practice. The new statistical criterion under the same conditions reduces the probability of errors to 0,075, which is already quite acceptable for a number of applications of neural network biometrics. By multiplying the two new criteria, it is possible to reduce the error probabilities to a probability of 0,045.

© Перфилов К. А., Полковникова С. А., Малыгин А. Ю., Куприянов Е. Н., 2022

Keywords: statistical analysis of small samples, normality hypothesis testing, geometric mean tests, differentiation of random data of small samples

Современная математическая статистика в том виде, какой мы ее знаем, опирается на хи-квадрат критерий Пирсона, построенный им в 1900 году. Эта математическая конструкция оказалась популярной из-за того, что она проста и долгое время альтернативы ей не было равных по эффективности [1, 2]. К сожалению, на малых выборках хи-квадрат критерий плохо работает. В 20 веке было разработано более 21 статистического критерия, которые на данный момент стали классикой [2, 3]. В этом веке, так же идет активная разработка новых статистических критериев среднего геометрического [4–6], ориентированных на снижения вероятностей ошибок первого и второго рода.

Численное моделирование критерия среднего геометрического может быть выполнено программой на языке MathCAD, листинг которой приведен на рис. 1.

$$
\begin{aligned}
sx(rr) := \;& x \leftarrow sort(rnorm(16,0,1+rr)) \\
& m \leftarrow mean(x) \\
& \sigma \leftarrow stdev(x) \\
& d \leftarrow \sum_{i=0}^{14} \frac{(x_{i+1}-x_i)\cdot dnorm(x_i,m,\sigma)}{x_{15}-x_0} \\
& x \leftarrow \frac{x}{stdev(x)} \\
& x \leftarrow x - x_0 + 1 \\
& SG \leftarrow \left(\prod_{i=0}^{15} x_i\right)^{\frac{1}{15}} \\
& (d \;\; SG \;\; d\cdot SG)^T
\end{aligned}
\qquad
\begin{aligned}
sxr(rr) := \;& x \leftarrow sort(runif(16,-3-rr,3+rr)) \\
& m \leftarrow mean(x) \\
& \sigma \leftarrow stdev(x) \\
& d \leftarrow \sum_{i=0}^{14} \frac{(x_{i+1}-x_i)\cdot dnorm(x_i,m,\sigma)}{x_{15}-x_0} \\
& x \leftarrow \frac{x}{stdev(x)} \\
& x \leftarrow x - x_0 + 1 \\
& SG \leftarrow \left(\prod_{i=0}^{15} x_i\right)^{\frac{1}{15}} \\
& (d \;\; SG \;\; d\cdot SG)^T
\end{aligned}
$$

Рис. 1. Листинг программы, позволяющей воспроизводить новый статистический критерий среднего гармонического SG

Результаты численного моделирования критерия среднего геометрического для малых выборок в 16 опытов приведены на рис. 2.

Из рисунка видно, что статистический критерий среднего геометрического имеет значение вероятностей ошибок первого и второго рода велики $P_1 = P_2 = P_{EE} = 0{,}141$, что примерно в 2,1 раза меньше ошибок хи-квадрат критерия. Существенно снизить вероятности

ошибок удается, если воспользоваться дифференциальным статистическим критерием – D. При этом вероятности ошибок снижаются еще в два раза до величины $P_1 = P_2 = P_{EE} = 0{,}076$. На рис. 3 отображены распределения выходных состояний дифференциально-разностного критерия $-D$.

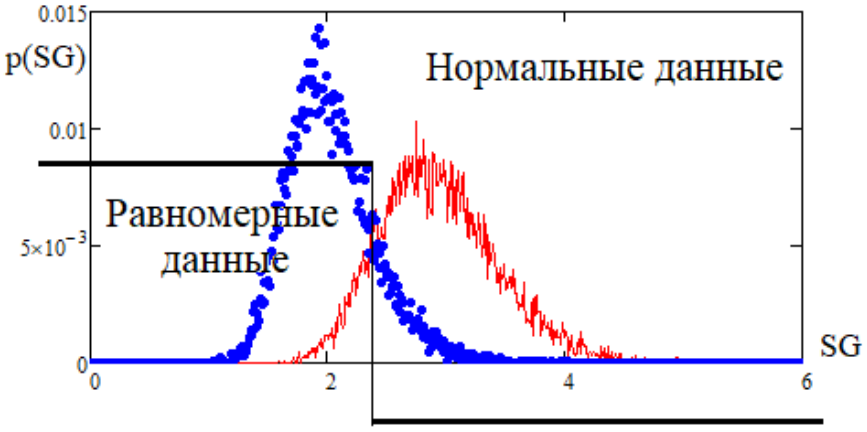

Рис. 2. Распределение среднего геометрического для малых выборок в 16 опытов

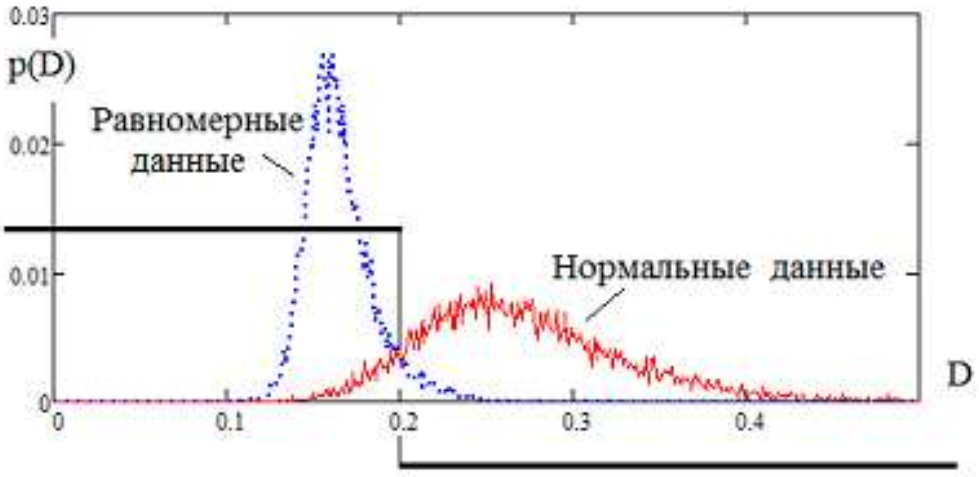

Рис. 3. Отклики дифференциально-разностного критерия при воздействии на него выборками в 16 опытов

Следует отметить, что синтез новых статистических критериев является достаточно редким событием. Так в прошлом веке были созданы примерно 21 критерий проверки гипотезы нормальности и порядка 200 критериев, ориентированных на проверку иных гипотез [2]. То есть создавалось порядка 2-х критериев в год. В этом веке за первые 20 лет создано менее 5 новых статистических критериев по проверке гипотезы нормальности.

Интересно отметить то, что работы по синтезу новых статистических критериев показали новые перспективные возможности. В частности, произведение состояний уже известных и новых статистических критериев дает новые более эффективные статистические критерии. Появляется возможность формального синтеза множества новых статистических критериев, перемножая парами множества уже известных статистических критериев. Так же можно создавать новые критерии перемножая их тройки и группы статистических критериев иных размеров. В качестве примера мы можем рассмотреть произведение двух, рассмотренных в данной работе критериев. Подобный численный эксперимент может быть проведен, опираясь на программное обеспечение, приведенное на рис. 1. Результаты такого численного эксперимента отражены на рис. 4.

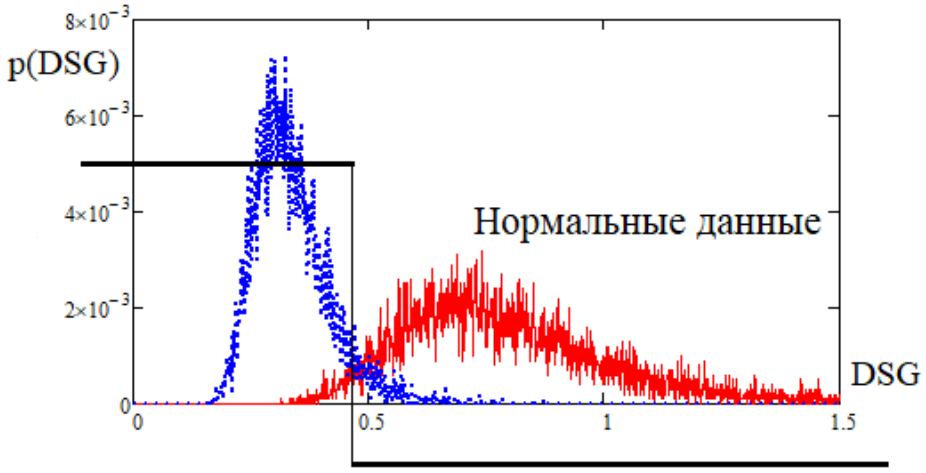

Рис. 4. Отклики произведения дифференциально-разностного критерия и критерия среднего геометрического при воздействии на них одной и той же выборкой в 16 опытов

Сравнивая данные, приведенные на рис. 3, 4, мы видим значительный рост мощности нового критерия по сравнению с лучшим из двух мультипликативно, объединяемых критериев. Третий новый критерий, полученный перемножением данных, дает равные вероятности ошибок первого и второго рода на уровне $P_1 = P_2 = P_{EE} = 0{,}045$.

Список литературы

1. Р 50.1.037–2002 Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Ч. I. Критерии типа χ2. М. : Госстандарт России, 2001. 140 с.

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

3. Р 50.1.037–2002. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Часть II. Непараметрические критерии. М. : Госстандарт России, 2002. 123 с.

4. Иванов А. И., Перфилов К. А., Малыгина Е. А. Оценка качества малых выборок биометрических данных с использованием дифференциального варианта статистического критерия среднего геометрического // Вестник Сибирского государственного аэрокосмического университета имени акад. М. Ф. Решетнева. 2016. № 4 (17). С. 864–871.

5. Иванов А. И., Банных А. Г., Куприянов Е. Н. [и др.]. Коллекция искусственных нейронов эквивалентных статистическим критериям для их совместного применения при проверке гипотезы нормальности малых выборок биометрических данных // Безопасность информационных технологий : сб. науч. ст. по материалам I Всерос. науч.-техн. конф. (г. Пенза, 24 апреля 2019 г.). Пенза, 2019. С. 156–164.

6. Иванов А. И. Искусственные математические молекулы: повышение точности статистических оценок на малых выборках (программы на языке MathCAD) : препринт. Пенза : Изд-во ПГУ, 2020. 36 с.

Для цитирования: Перфилов К. А., Полковникова С. А., Малыгин А. Ю., Куприянов Е. Н. Мультикативное объединение двух новых статистических критериев для их взаимного усиления при проверке гипотезы нормальности данных малой выборки // Безопасность информационных технологий : сб. науч. ст. по материалам IV Всерос. науч.-техн. конф. : в 2 т. Пенза : Изд-во ПГУ, 2022. Т. 1. С. 49–53.

УДК 519.23

КАЛЬКУЛЯТОР ОЦЕНКИ ЭНТРОПИИ КОДОВ ОТКЛИКОВ НЕЙРОСЕТИ НА ПРИМЕРЕ РУКОПИСНОГО ОБРАЗА «ЧУЖОЙ» В ПРОСТРАНСТВЕ КОРРЕЛЯЦИОННЫХ СВЯЗЕЙ МЕЖДУ ИХ РАЗРЯДАМИ

В. А. Сумин

Пензенский государственный университет, г. Пенза

Аннотация. Исследуются принципы работы биометрико-нейросетевой аутентификации личности путем преобразования личных биометрических данных человека в его криптографический ключ. Параллельно с анализом кодов в пространстве расстояний Хэмминга исследуется возможность оценки энтропии в пространстве корреляционных связей разрядов, анализируемых кодов.

Ключевые слова: биометрическая аутентификация, биометрический образ «Свой», биометрический образ «Чужой», коэффициенты корреляции между разрядами кода, расстояние Хемминга

ENTROPY ESTIMATION CALCULATOR OF NEURAL NETWORK RESPONSE CODES ON THE EXAMPLE OF THE «ALIEN'S» HANDWRITTEN IMAGE IN THE SPACE OF CORRELATION BETWEEN THEIR DIGITS

V. A. Sumin

Penza State University, Penza

Abstract. The principles of operation of biometric-neural network authentication of a person are investigated by converting personal biometric data of a person into his cryptographic key. In parallel with the analysis of codes in the space of Hamming distances, the possibility of estimating the entropy in the space of correlation links of digits of the analyzed codes is investigated.

Keywords: biometric authentication, «Own» biometric image, «Alien» biometric image, correlation coefficients between code digits, Hamming distance

Больше половины мирового населения имеют аккаунты на одной или нескольких платформах. Как показывает практика, парольная защита доступа к личным интернет-ресурсам имеет ряд уязвимостей. Пользователи обычно применяют короткие, легко подбираемыми пароли. Для повышения безопасности парольной защиты в настоящее время желательно использовать биометрическую аутентификацию.

© Сумин В. А., 2022

Биометрическая аутентификация – процесс доказательства и проверки подлинности заявленного пользователем имени, через предъявление пользователем своего биометрического образа «Свой» и путём преобразования этого образа в соответствии с заранее определённым протоколом нейросетевой аутентификации по требованиям ГОСТ Р 52633.0–2006 [1].

Корректная работа биометрической аутентификации заключается в автоматическом обучении нейросетевого преобразователя по ГОСТ Р 52633.5–2011 [2]. Для обучения нейросетевого преобразователя биометрический код использует от 8 до 20 примеров образа «Свой» и не менее 256 примеров образа «Чужой».

Нейросетевой преобразователь – преобразователь, заранее обученный преобразовывать многомерные континуумы примеров биометрического образа «Свой» в почти однозначный выходной код криптографического ключа или длинного пароля доступа по ГОСТ Р 52633.0–2006. После каждого обучения нуждается в быстром автоматическом тестировании.

Биометрический образ «Свой» – биометрический образ человека, параметры которого порождают с высокой вероятностью от 0,92 до 0,95 на выходах обученной нейронной сети код его личного криптографического ключа.

Биометрический образ «Чужой» – случайный биометрический образ, параметры которого порождают на выходах обученной нейронной сети случайный выходной код.

Так как биометрические образы могут быть подвержены к атакам подбора, основным требованием является вычисление вероятности ошибок второго рода и выполнение процедур тестирования.

Ошибка второго рода – вероятность ложной идентификации пользователя, отсутствующего в базе данных. Такая ошибка должна оцениваться либо по международным рекомендациям стандарта ГОСТ Р ИСО/МЭК 19795-1–2007, либо ГОСТ З 52633.3–2011.

Если производить оценку по ГОСТ Р ИСО/МЭК 19795-1–2007, то придется использовать тестовые базы с объемом тестовых образов «Чужой» в 30 раз больше, чем обратная величина предполагаемой стойкости к атакам подбора. Так, если будет тестироваться защита образа с вероятностью ошибок в 0,0000001, размер тестовой базы должен составлять боле 30 миллионов образов «Чужой». Это является значительным юридическим барьером для выполнения лабораторных работ по биометрии для университетов. Во всех странах собор, хранение и использовать большие базы биометрических образов ЗАПРЕЩЕНО. Именно по этой причине мы постоянно даем «согласие» на обработку своих персональных данных.

Решить проблему возможно, если воспользоваться отечественным стандартом по быстрому тестированию на малых выборках. ГОСТ Р 52633.3–2011. Этот стандарт требует перехода от анализа обычных кодов к расстояниям Хэмминга между ними.

Метрика Хэмминга позволяет значительно сэкономить вычислительные ресурсы при реализации генетических алгоритмов направленного перебора. В рамках гипотезы нормального распределения расстояний Хэмминга позволяет вычислить математическое ожидание и стандартное отклонение [5]. По этим параметрам можно вычислить вероятность ошибок второго рода, что позволит определить процентное прохождение аутентификации биометрического образа «Чужой».

Немаловажно произвести логарифмические вычисление для поиска энтропии образа «Чужой». По результатам вычисления минимальная энтропия образа «Чужой» показывает, насколько ближе к образу «Свой». Особенностью данных расчетов является оценка расстояний Хэмминга, которая в свою очередь требует данные в виде примеров образа «Свой» и данные о биокоде «Свой».

Корреляционная метрика среднего значения модулей коэффициентов парной корреляции способствует вычислению энтропии, минуя промежуточную оценку расстояний Хэмминга [6]. Каждый из образов будет иметь свое распределение корреляции, на основе которых будет сделан вывод о стойкость к атакам подбора.

В работе [7] путем проведения численного эксперимента было показано, что шкала расстояний Хэмминга, используемая ГОСТ Р 52633.3 [4] и шкала взаимных коэффициентов корреляции, исследуемых разрядов кодов обратны. К сожалению, численный эксперимент с был проведен на минимальной выборке в два образа «Чужой». Столь малая выборка была обусловлена необходимостью ручного редактирования отчетного файла среды моделирования «БиоНейроАвтограф» [6] testKeys.txt. Экранная форма стандартного редактора отчетного файла testKeys.txt, приведена на рис. 1.

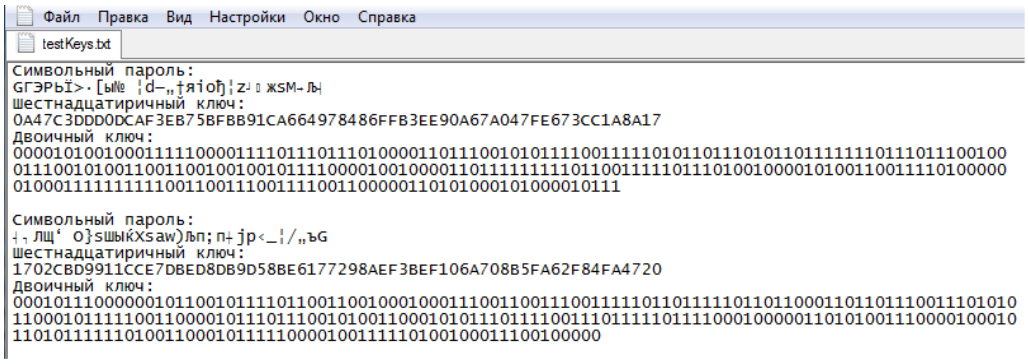

Рис. 1. Экранная форма стандартного редактора, которую приходилось при ручном редактировании, так как формат файла testKeys.txt специфичен и не читается обычными средствами автоматизации вычислений

Автоматизация чтения отчетного файла testKeys.txt среды моделирования «БиоНейроАвтограф» выполняется при запуске, разработанного на C++ калькулятора коэффициентов корреляции. Экранная форма, разработанного калькулятора приведена на рис. 2.

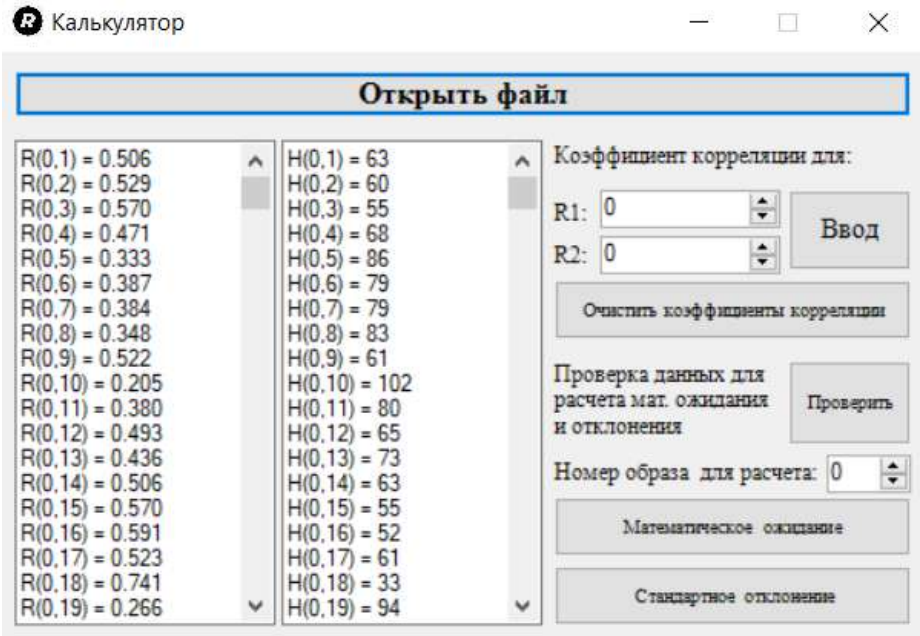

Рис. 2. Экранная форма, разработанного корреляционного калькулятора

Для того, чтобы результаты вычислений «корреляционного калькулятора» были корректны перед его использованием необходимо удалять (стирать) уже существующий файл в среде моделирования DATA/testKeys.txt. Далее следует запусти среду моделирования в режиме тестирования для примеров одного биометрического образа «Чужой» [5]. При этом среда моделирования формирует новый файл отчетности DATA/testKeys.txt.

Программное обеспечение читает этот новый файл DATA/test Keys.txt, получая из него, например, 20 бинарных ключей длинной в 256 бит. Далее калькулятор вычисляет (20×20 – 20)/2 = 190 коэффициентов корреляции и расстояний Хэмминга. Очевидно, что 190 случайных коэффициентов корреляции и 190 расстояний Хэмминга должны давать более достоверные статистические оценки в сравнении с алгоритмом предсказания по ГОСТ Р 52633.3 [4], использующем только 20 расстояний Хэмминга [8].

В силу того, что корреляционный калькулятор должен давать доступ к большим массивам данных в его экранной форме предусмотрено два скроллинга для отображения массива коэффициентов корреляции и массива расстояний Хэмминга с отображением трех значащих десятичных цифр (рис. 2). Для просмотра точного значения вычисляемых параметров (6 десятичных цифр) предусмотрен ручной

набор номеров примеров образа «Чужой». Параллельно с выводом на экран корреляционный коррелятор записывает в данные, постоянно обновляемые в отчетные файлы DATA/testKeys_R.txt и DATA/test Keys_H.txt.

Список литературы

1. ГОСТ Р 52633.0–2006. Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации.

2. ГОСТ Р 52633.5–2011. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа.

3. ГОСТ Р ИСО/МЭК 19795-1–2007. Автоматическая идентификация. Идентификация биометрическая. Эксплуатационные испытания и протоколы испытаний в биометрии. Ч. 1. Принципы и структура.

4. ГОСТ Р 52633.3–2011. Защита информации. Техника защиты информации. Тестирование стойкости средств высоконадежной биометрической защиты к атакам подбора.

5. Иванов А. И. Тестирование нейронных сетей, обученных алгоритмом ГОСТ Р 52633.5–2022, в среде моделирования «БиоНейроАвтограф» : учеб.-метод. пособие. Пенза : Изд-во ПГУ, 2020. 36 с. URL: https://tsib.pnzgu.ru/page/39329

6. Иванов А. И., Захаров О. С. Среда моделирования «БиоНейроАвтограф». URL: http://пниэи.рф/activity/science/noc/bioneuroautograph.zip

7. Горбунов К. А., Никитин В. В. Нейросетевая биометрия: подтверждение гипотезы обратных шкал для метрики корреляционной сцепленности и метрики расстояний хэмминга при их применении к ключам-откликам на примеры одного образа «Чужой» // Безопасность информационных технологий : сб. науч. ст. по материалам III Всерос. науч. техн. конф. Пенза : Изд-во ПГУ, 2021. Т. 1. С. 83–86.

8. Иванов А. И., Иванов А. П., Ратников К. А. Статистико-нейросетевой анализ биометрических образов в пространствах спектров кроссверток и автосверток Хэмминга : препринт. Пенза : Изд-во ПГУ, 2021. 56 с. doi:10.13140/RG.2.2.16514.35524

Для цитирования: Сумин В. А. Калькулятор оценки энтропии кодов откликов нейросети на примере рукописного образа «Чужой» в пространстве корреляционных связей между их разрядами // Безопасность информационных технологий : сб. науч. ст. по материалам IV Всерос. науч.-техн. конф. : в 2 т. Пенза : Изд-во ПГУ, 2022. Т. 1. С. 54–58.

УДК 519.24;53;57.017

СИНТЕЗ НОВОГО СТАТИСТИЧЕСКОГО КРИТЕРИЯ ДЛЯ ПРОВЕРКИ ГИПОТЕЗЫ НЕЗАВИСИМОСТИ, СЛАБО СВЯЗАННОГО С ОЦЕНКАМИ ПО КЛАССИЧЕСКОЙ ФОРМУЛЕ ЭДЖУОРТА – ЭУДЛТОНА – ПИРОНА

Ю. И. Серикова

Пензенский государственный университет, г. Пенза

Аннотация. Рассматривается проблема оценки коэффициентов корреляции, дополняющих классическую формулу Эджуорта – Эудлтона – Пирсона 1890 г. Показано, что дискретный вариант классического критерия дает сильно коррелированные результаты с его континуальным аналогом. Разрушить корреляцонную связь данных удается в случае, если воспользоваться проверкой близости числа точек, попадающих в различные квадранты декартовой системы координат.

Ключевые слова: критерии проверки гипотезы независимости, неросетевой эквивалент критерия проверки, устранение связи выходных состояний двух и более статистических критериев

SYNTHESIS OF A NEW STATISTICAL CRITERION FOR TESTING THE HYPOTHESIS OF INDEPENDENCE WEAKLY RELATED TO EVALUATIONS FROM THE CLASSICAL EDGEWORTH – AUDLETON – PYRON FORMULA

Yu. I. Serikova

Penza State University, Penza

Abstract. The problem of estimating the correlation coefficients supplementing the classical Edgeworth – Eudleton – Pearson formula of 1890 is considered. It is shown that the discrete version of the classical criterion gives strongly correlated results with its continual counterpart. It is possible to destroy the correlation connection of data if we use the check of the proximity of the number of points that fall into different quadrants of the Cartesian coordinate system.

Keywords: criteria for testing the hypothesis of independence, non-network equivalent of the test criterion, elimination of the connection between the output states of two or more statistical criteria

В настоящее время известно порядка 21 статистических критериев проверки гипотезы независимости [1]. Каждому критерию

© Серикова Ю. И., 2022

можно поставить в соответствие эквивалентный ему искусственный нейрон. Это позволяет при проверках той или иной статистической гипотезы использовать несколько критериев или эквивалентных им искусственных нейронов.

При параллельном использовании нескольких критериев (искусственных нейронов) важно пытаться обеспечить отсутствие у их откликов корреляционных связей. Это далеко не всегда возможно. Так мы можем упорядочить данные по одной из координат и оценивать корреляционные связи – sxr как отношение длин случайно составляющей к монотонно возрастающей детерминированной составляющей [2, 3]. При этом мы получаем оценки со слабой зависимостью с данными, полученными по классической формуле.

Если использовать подсчет числа точек, анализируемых выборок, попавших в один из четырех квадрантов декартовой системы координат, то мы получим дискретный аналог классической формулы оценки корреляции Эджуорта – Эудлтона – Пирсона [4]. Пример подобного анализа иллюстрирует рис. 1.

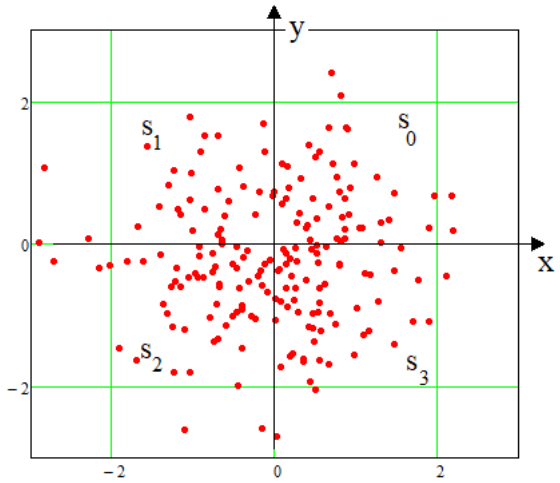

Рис. 1. Случайное расположение центрированной и нормированной выборки в 200 точек $\{x_i, y_i\}$ при использовании системы декартовых координат

Программное обеспечение такого анализа данных и результаты численного эксперимента приведены на рис. 2.

Из рис. 2 видно, что, рассматриваемый статистический критерий является дискретным. Выходные данные всегда дискретны, в этом контексте эту математическую конструкцию можно рассматривать как некоторую математическую молекулу. Такая молекула всегда имеет линейчатый спектр своих выходных состояний. Линии допустимых выходных состояний такой молекулы для выборки в 200 опытов расположены с шагом 0,01. Изменение размеров

обрабатываемой выборки приводит к изменению шага линий выходного спектра. Рост выборки приводит к снижению равномерного шага между линиями.

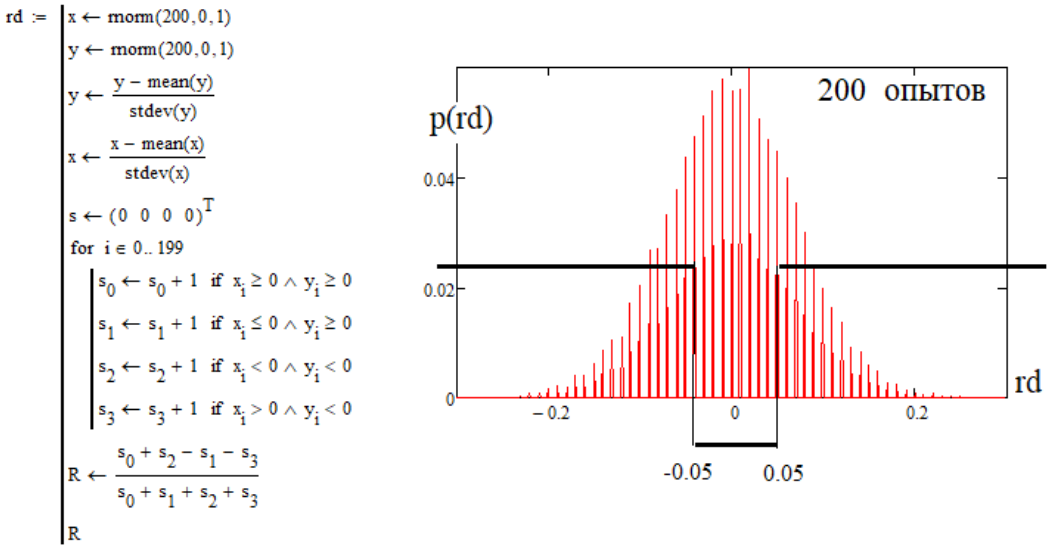

Рис. 2. Дискретный статистический критерий корреляции со спектром линий, расположенных с шагом $\Delta rd = 0{,}01$ между ними

К сожалению, рассматриваемая математическая конструкция имеет сильную корреляционную связь с классической формулой Эджуорта – Эудлтона – Пирсона:

$$\begin{cases} corr(rd, r) = 0{,}641 \\ corr(rd, mmrd) = 0{,}00 \\ corr(r, mmrd) = 0{,}00\ . \\ corr(r, sxr) = 0{,}00 \\ corr(rd, sxr) = 0{,}00 \end{cases} \qquad (1)$$

Следует отметить, что такую математическую конструкцию можно рассматривать как два персептрона, выполняющие разделение плоскости двумя перпендикулярными линиями, анализируемых данных. Такие персептроны откликаются разными знаками ± на входные данные, находящиеся в разных квадрантах декартовой системы координат.

Разрушить сильную корреляционную связь классической формулы и критерия – *rd* удается в случае перехода к анализу вариаций сочетаний чисел точек, обнаруженных в каждом квадранте. Программное обеспечение численного эксперимента и его результаты отображены на рис. 3.

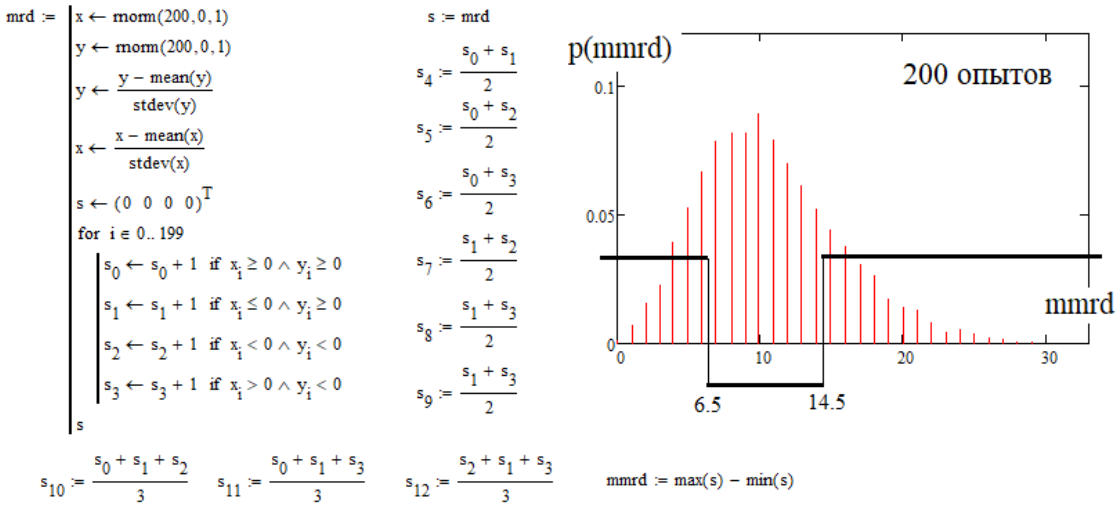

Рис. 3. Второй дискретный статистический критерий корреляции со спектром линий, расположенных с шагом $\Delta mmrd = 1{,}00$ между ними

Операции по поиску минимальных и максимальных значений вариаций данных разрушают корреляционные связи, т.е. коэффициенты корреляции оказываются нулевыми (1).

Список литературы

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

2. Волчихин В. И., Иванов А. И., Ахметов Б. Б., Серикова Ю. И. Фрактально-корреляционный функционал, используемый при поиске пар слабо зависимых биометрических данных в малых выборках // Вестник высших учебных заведений. Поволжский регион. Технические науки. 2016. № 4. С. 25–31.

3. Волчихин В. И., Иванов А. И., Сериков А. В., Серикова Ю. И. Тестирование аналогового и квантового оракулов линейной вычислительной сложности, предсказывающих значения коэффициента корреляции на малой выборке в 32 опыта // Известия высших учебных заведений. Поволжский регион. Технические науки. 2017. № 3. С. 70–80.

4. Волчихин В. И., Иванов А. И., Сериков А. В., Серикова Ю. И. Квантовая суперпозиция дискретного спектра состояний математической молекулы корреляции для малых выборок биометрических данных // Вестник Мордовского университета. 2017. Т. 27, № 2. С. 230–243.

Для цитирования: Серикова Ю. И. Синтез нового статистического критерия для проверки гипотезы независимости, слабо связанного с оценками по классической формуле Эджуорта – Эудлтона – Пирона // Безопасность информационных технологий : сб. науч. ст. по материалам IV Всерос. науч.-техн. конф. : в 2 т. Пенза : Изд-во ПГУ, 2022. Т. 1. С. 59–62.

УДК 004.056

ЛИНЕЙНАЯ СВЯЗЬ СТАНДАРТНОГО ОТКЛОНЕНИЯ ОШИБКИ ВЫЧИСЛЕНИЯ КОЭФФИЦИЕНТОВ КОРРЕЛЯЦИИ С ОБЪЕМОМ ВЫБОРКИ В ЛОГАРИФМИЧЕСКИХ КООРДИНАТАХ

Т. А. Золотарева[1], С. В. Качалин[2], А. С. Боровский[3]

[1]Липецкий государственный педагогический университет имени П. П. Семенова-Тян-Шанского, г. Липецк
[2]Научно-производственное предприятие «Рубин», г. Пенза
[3]Оренбургский государственный университет, г. Оренбург

Аннотация. Рассмотрена линейная связь стандартного отклонения ошибок вычисления малых значений коэффициентов корреляции по формуле Пирсона – Эджуорта – Эдултона с объемом выборки в логарифмических координатах.

Ключевые слова: коэффициент корреляции, нейросетевые преобразователи, выборка данных, квадратичные нейроны

LINEAR RELATIONSHIP OF THE STANDARD DEVIATION OF THE ERRORS OF CALCULATION OF THE CORRELATION COEFFICIENTS WITH THE SAMPLE VOLUME IN LOGARITHMIC COORDINATES

T. A. Zolotareva[1], S. V. Kachalin[2], A. S. Borovsky[3]

[1]Lipetsk State Pedagogical University named after P. P. Semenov-Tyan-Shansky, Lipetsk
[2]Research and Production Enterprise «Rubin», Penza
[3]Orenburg State University, Orenburg

Abstract. The linear relationship between the standard deviation of errors in calculating small values of correlation coefficients using the Pearson – Edgeworth – Edulton formula and the sample size in logarithmic coordinates is considered.

Keywords: correlation coefficient, neural network converters, data sampling, quadratic neurons

На текущий момент нейросетевые преобразователи биометрий в код криптографического ключа обогащают «сырые» данные в линейном пространстве и обучаются автоматически по алгоритму ГОСТ Р 52633.5–2011 на обучающих выборках объемом от 16 до 20

© Золотарева Т. А., Качалин С. В., Боровский А. С., 2022

примеров образа «Свой». Предположительно в ближайшем будущем в России будет разработан стандарт, регламентирующие следующий алгоритм обучения квадратичных нейронов с многоуровневыми квантователями [1].

Из-за наметившегося перехода к использованию квадратичных нейронов возникает интерес к задаче регуляризации обучения нейронов Махалонобиса [2] и нейронов Байеса [3 4]. Проблема состоит в том, что коэффициенты корреляции между двумя параметрами на малых выборках по формуле Пирсона плохо обусловлены. Как результат необходимо выполнить некоторую нейросетевую регуляризацию статистических вычислений, например, путем замены нескольких статистических критериев их эквивалентными искусственными нейронами [5].

В связи с этим, скорее всего в ближайшем будущем должны быть созданы программные калькуляторы, которые должны быть способны выполнять регуляризацию задачи вычисления коэффициентов корреляции уменьшая ошибку их вычисления на малых выборках [6].

Очевидно, что экранные формы и интерфейс управления калькулятором вычисления коэффициентов корреляции будет слабо зависеть от того, сколько нейронных сетей будет в нем использовано. Очевидно, так же, что интерфейс регуляризатора вычислений должен обеспечивать ввод данных из внешнего файла «rr.txt» и редактирование этих данных.

Важнейшим элементом обеспечения доверия к новому классу программ (нейросетевым корректорам ошибок вычисления коэффициентов корреляции) является наличие встроенного режима тестирования.

Оценку коэффициента повышения точности вычислений калькулятора с нейросетевой регуляризацией следует оценивать как отношение стандартных отклонений:

$$k = \frac{\sigma(r)}{\sigma(rr)}. \qquad (1)$$

При этом выигрыш по эквивалентному увеличению объема выборки исходных данных следует вычислять как квадрат отношения (1). Проверить корректность этих вычислений, можно воспользовавшись любой из стандартных математических программ. Для этого достаточно сформировать нужную тестовую выборку и вычислить по ней значения по классической формуле Эджиорта – Эудлтона – Пирсона и по нейросетевым формулам калькулятора. Отношение

стандартных отклонений (1) даст нам возможность оценить повышение точности калькулятора в сравнении с классической формулой.

В конце XIX века уже использовалась при вычислении парных коэффициентов корреляции классическая формула Эджуорта – Эудлтона – Пирсона:

$$r(x,y)=\frac{1}{n}\sum_{i=1}^{n}\frac{\left(E(x)-x_i\right)\cdot\left(E(y)-y_i\right)}{\sigma(x)\cdot\sigma(y)}, \qquad (2)$$

где $E(.)$ – математические ожидания; $\sigma(.)$ – стандартные отклонения, сравниваемых выборок; n – размер выборок.

Еще одной проблемой решаемой задачи является то, что распределения ошибок оценки коэффициентов корреляции оказываются симметричными только в одной точке $E(r) = 0{,}0$. При других значениях математических ожиданий коэффициентов корреляции, распределения ошибок оказываются асимметричными.

В случаях, когда объем памяти процессора ограничен (например, когда в качестве доверенного вычислителя используется SIM-карта) необходимо от табличных вычислений переходить к аналитической форме описания интервалов неопределенности. В случае симметричных интервалов неопределенности мы можем воспользоваться инженерным правилом трех стандартных отклонений:

$$\Delta r=\pm 3\cdot\sigma(r), \qquad (3)$$

Следует так же отметить, что в логарифмических координатах стандартное отклонение связано с числом опытов линейной функцией.

В итоге мы получаем простую функциональную связь в логарифмических координатах:

$$\log_{10}(\sigma(r))=1-0{,}5\cdot\log_{10}(n). \qquad (4)$$

Объединив выражения (3) и (4), мы всегда можем вычислить значение стандартного отклонения для выборок любого объема и оценить симметричные интервалы неопределенности.

Таким образом, можно говорить о линейной связи стандартного отклонения ошибок вычисления малых значений коэффициентов корреляции по формуле Пирсона – Эджуорта –Эдултона с объемом выборки в логарифмических координатах.

Список литературы

1. Иванов А. И., Безяев А. В., Малыгина Е. А., Серикова Ю. И. Второй национальный стандарт России по быстрому автоматическому обуче-

нию больших искусственных нейронных сетей на малых выборках биометрических данных // Безопасность информационных технологий : сб. науч. ст. по материалам I Всерос. науч.-техн. конф. (г. Пенза, 24 апреля 2019 г.). Пенза, 2019. С. 174–177.

2. Серикова Ю. И. Двойная регуляризация процедур обучения нейронов Махаланобиса за счет симметризации корреляционных связей и компенсации ошибок вычисления коэффициентов парной корреляции биометрических данных // Безопасность информационных технологий : сб. науч. ст. по материалам I Всерос. науч.-техн. конф. (г. Пенза, 24 апреля 2019 г.). Пенза, 2019. С. 26–34.

3. Ложников П. С. Биометрическая защита гибридного документооборота : монография. Новосибирск : Изд-во СО РАН, 2017. 130 с.

4. Ложников П. С. Методология защиты смешанного документооборота на основе многофакторной биометрической аутентификации с применением нейросетевых алгоритмов : автореф. дис. … д-ра техн. наук. 05.13.19. Уфа : Уфимский гос. авиац. техн. ун-т, 2019. 32 с.

5. Иванов А. И., Золотарева Т. А. Искусственный интеллект в защищенном исполнении: синтез статистико-нейросетевых автоматов многокритериальной проверки гипотезы независимости малых выборок биометрических данных : препринт. Пенза : Изд-во ПГУ, 2020. 105 с.

6. Качалин С. В., Савинов К. Н., Иванова Н. А., Золотарева Т. А. Минимальный функционал калькулятора, выполняющего нейросетевую регуляризацию вычисления коэффициентов корреляции на малых выборках биометрических данных // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. Пенза : Изд-во ПГУ, 2020. С. 37–41.

Для цитирования: Золотарева Т. А., Качалин С. В., Боровский А. С. Линейная связь стандартного отклонения ошибки вычисления коэффициентов корреляции с объемом выборки в логарифмических координатах // Безопасность информационных технологий : сб. науч. ст. по материалам IV Всерос. науч.-техн. конф. : в 2 т. Пенза : Изд-во ПГУ, 2022. Т. 1. С. 63–66.

УДК 519.24;53;57.017

ОБЗОР ПРИЗНАКОВ, ИЗВЛЕКАЕМЫХ ИЗ РЕЧЕВЫХ СИГНАЛОВ С ЦЕЛЬЮ РАСПОЗНАВАНИЯ СОСТЯЗАТЕЛЬНЫХ ПРИМЕРОВ

И. Е. Панфилова[1], А. Е. Сулавко[2]

[1]Самарский государственный технический университет, г. Самара
[2]Омский государственный технический университет, г. Омск

Аннотация. Для защиты систем верификации диктора от нового класса угроз под названием «дипфейк» из речевого сигнала, которым оперирует система, необходимо извлекать признаки, наиболее полно отражающие различия между подлинной речью человека и синтетической подделкой. Представлен обзор таких характеристик речевого сигнала.

Ключевые слова: верификация диктора, дипфейк, голосовая биометрия

REVIEW OF FEATURES EXTRACTED FROM SPEECH SIGNALS FOR THE PURPOSE OF RECOGNITION OF ADVERSAL EXAMPLES

I. E. Panfilova[1], A. E. Sulavko[2]

[1]Samara State Technical University, Samara
[2]Omsk State Technical University, Omsk

Abstract. To protect speaker verification systems from a new class of threats called «deepfake», it is necessary to extract features from the speech signal that the system operates that most fully reflect the differences between genuine human speech and a synthetic fake. The article presents an overview of such characteristics of a speech signal.

Keywords: speaker verification, deepfake, voice biometrics

Введение

Применение биометрических систем идентификации/верификации пользователя на сегодняшний день является стандартной практикой во множестве областей человеческой деятельности, в связи с чем, вопросы обеспечения безопасности таких систем (в том числе их устойчивости к разного рода атакам), становятся все более актуальными и представляют особый исследовательский интерес. Одной из технологий, способной внести колоссальные изменения в представления

© Панфилова И. Е., Сулавко А. Е., 2022

о надежности биометрических систем, являются дипфейки (конкатенация слов «глубинное обучение» (англ. *deeplearning*) и «подделка» (англ. *fake*)). В сущности, дипфейк можно представить в качестве частного случая спуфинга (англ. *spoofing* – подмена) или состязательной атаки, представляющей собой атаку, при которой один человек или программа успешно маскируется под другую путём фальсификации данных. В случае дипфейков генерация синтетических данных, как правило, осуществляется с использованием методов глубокого обучения, таких как генеративно-состязательные сети (GAN) и рекуррентные нейронные сети (RNN) [1]. Сильным толчком к развитию указанной технологии во многом послужил бурный рост объемов данных, представленных в открытом доступе в сети Интернет, что позволило без труда генерировать достаточно большие датасеты для обучения таких нейронных сетей.

Сегодня под термином «дипфейк» чаще всего понимают методику синтеза изображения с целью замены определённых его элементов на желаемые образы. Как правило, это наложение изображения лица одного человека (целевого объекта) на видео с другим человеком (исходным объектом) для получения видео, на котором целевой объект делает или говорит что-то, что делает или говорит исходный объект [2]. Однако ошибочно полагать, что дипфейки используются исключительно с целью подделки изображений. Одной из систем, не менее сильно подверженных данной атаке, являются системы автоматической верификации диктора [3] (automatic verification speaker – ASV, АВД). В данном случае спуфинг осуществляется по средствам так называемых *голосовых дипфейков*, призванных за счет использования методов синтезирования речи и/или ее преобразования заставить систему верификации принять искусственно сгенерированную речь за подлинную (принадлежащую пользователю системы).

Учитывая достижения последних лет в области синтезирования искусственной речи, а также методы преобразования голоса с целью произнесения текста (парольной фразы) голосом другого человека, важно исследовать и определять различия между подлинной и искусственно синтезированной речью. Во многом такие различия заключаются в признаках, извлекаемых из голосовых сигналов и в их дальнейшей обработке системой верификации диктора. Приведение обзора наиболее часто встречающихся на сегодняшний день признаков голосового сигнала, используемых в системах автоматической верификации диктора и позволяющих детектировать наличие синтетической речи (дипфейка), и является целью данной статьи.

Верификация диктора на основе изображений

Подход, представляющий собой верификацию диктора на основе изображений, заключается в представлении звукового сигнала диктора в визуальной форме с последующим использованием полученных изображений для извлечения дополнительных признаков или для подачи на вход классификатора. Двумя наиболее информативными визуальными представлениями речевого сигнала на сегодняшний день являются *спектрограммы* и *мел-спектрограммы*. Спектрограмма представляет собой изображение, демонстрирующее зависимость спектральной плотности мощности сигнала от времени, т.е. показывает изменения частоты акустических сигналов во времени. Чаще всего такое изображение генерируется путем выполнения оконного преобразования Фурье:

$$F(t,\omega) = \int_{-\infty}^{\infty} f(\tau) W(\tau - t) e^{-i\omega\tau} d\tau, \quad (1)$$

где $W(\tau - t)$ – некоторая оконная функция.

Мел-спектограммы [14], в свою очередь, являются разновидностью спектограмм, в которых частота представлена по шкале мел. Перевод в шкалу мел осуществляется с помощью преобразования [4]:

$$f_{mel} = 2595 log_{10}\left(1 + \frac{f}{700}\right), \quad (2)$$

где f – частота в Гц; f_{mel} – частота по шкале Мел.

Чаще всего, изображения, полученные путем вышеуказанных преобразований, подаются на вход глубоких нейронных сетей (DNN) с целью извлечения дополнительных признаков, позволяющих однозначно определить подлинную или синтезированную речь. В качестве архитектур глубоких нейронных сетей для таких задач могут использоваться:

– сверточные нейронные сети (CNN) [10, 11], наиболее прочно зарекомендовавшие себя в задачах распознавания образов;

– сверточные нейронные сети доверия [3, 15], в качестве слоев которой применяется сверточная ограниченная машина Больцмана (CRBM);

– временные сверточные сети (TCN) [1], позволяющие объединить кодируемую CNN пространственно-временную низкоуровневую информацию и классификатор, оперирующий высокоуровневой временной информацией.

Речевые признаки

Как видно из предыдущего раздела, для работы с речевым сигналом и выделения из него характерных для конкретного человека параметров, достаточно часто используются спектрограммы, а именно спектральный анализ. Такое внимание спектру речевого сигнала человека уделяется не случайно. Дело в том, что спектр сигнала содержит в себе полезную информацию, позволяющую достаточно точно проводить различия между синтезированной и подлинной речью. Помимо огибающей спектра, которая сама по себе является крайне информативным параметром голоса человека, так как описывает форму его голосового тракта (а также отражает относительный вклад гармоник в общую энергию речевого сигнала [6]), одними из наиболее часто используемых параметров спектра также являются частота основного тона F_0 (Fundamental Frequency) и формантные частоты (Formant Frequencies) [3].

Частота основного тона, или просто основная частота отражает высоту голоса и определяется как самая низкая частота периодического сигнала. Поскольку колебание исходит от органической структуры, оно не является строго периодическим, но содержит значительные флуктуации. В частности, величина изменения длины периода и амплитуды основной частоты носят названия джиттера и шиммера [3] соответственно. F_0 отдельного говорящего зависит прежде всего от длины голосовых связок, которая, в свою очередь, коррелирует с общими размерами тела. Культурные и стилистические аспекты речи, естественно, также имеют большое значение.

Форманта – это концентрация акустической энергии вокруг определенной частоты речевой волны. Есть несколько формант, каждая на своей частоте, примерно по одной в каждой полосе 1000 Гц. Или, другими словами, форманты встречаются с интервалом примерно в 1000 Гц. Каждая форманта соответствует резонансу в голосовом тракте и создает пики в огибающей спектра. По формантным частотам можно достаточно точно определять диктора, так как графики спектра воспроизведения одного и того же звука от двух дикторов будут отличаться [6].

Кепстральный анализ

Одним из базовых подходов в работе с речевыми сигналами является вычисление кепстральных коэффициентов, позволяющих наиболее точно описать огибающую спектра речевого сигнала. В общем виде кепстр представляет собой функцию обратного преобразования Фурье от логарифма спектра мощности сигнала и может характеризоваться следующим выражением:

$$C_s(q) = \frac{1}{2\pi} \int_{-\infty}^{\infty} ln\left|S(\omega)\right|^2 e^{i\omega q} d\omega, \qquad (3)$$

где $S(\omega)$ – спектр входного сигнала.

В виду того, что кепстр достаточно точно описывает спектр сигнала в сжатом виде, использование кепстрального анализа для верификации диктора представляется удобным с точки зрения эффективности и быстродействия [6].

Наиболее часто используемыми кепстральными коэффициентами в задачах верификации диктора являются *мел-кепстральные частотные коэффициенты* (mel-frequency cepstral coefficients (MFCC)) [5, 7, 12]. Данные коэффициенты наиболее точно описывают огибающую спектра, характеризующую голосовой тракт человека. Процесс извлечения мел-кепстральных коэффициентов (МКК) можно кратко описать следующим образом.

Сначала исходный речевой сигнал разбивается на фреймы (от 20 до 40 мс). Во избежание эффекта утечек из-за разрывов на концах речевого сигнала при применении преобразования Фурье, каждый фрейм умножается на оконную функцию Хэмминга, после чего к получившимся результатам применяют дискретное преобразование Фурье для получения спектра сигнала. Получившиеся спектры показывают количество энергии, присутствующей в различных частотных диапазонах. Поскольку человеческий слух чувствителен к более низким частотам, используется представление, наиболее точно описывающее такой феномен: *мел-спектр*. Для этого спектр раскладывается на мел шкале, с использованием банка фильтров (т.е. набора перекрывающихся нелинейных банков мел-фильтров, полоса пропускания которых сужается на низких частотах и расширяется на более высоких частотах). Выполняется перемножение периодограмм полученных спектров на указанные фильтры. В результате перечисленных операций получаются энергии набора фильтров, которые после логарифмирования формируют некоторый набор коэффициентов, еще не являющихся МКК. Сами мел-кепстральные коэффициенты можно получить путем применения последующего дискретного косинусного преобразования.

В зависимости от банка фильтров, используемого для разложения спектра сигнала на мел шкале, получаемые коэффициенты могут выделять разную информацию из частотных характеристик сигнала. Так, помимо МКК, банк треугольных фильтров используется при вычислении линейно-частотных кепстральных коэффициентов (linear

frequency cepstral coefficients (LFCC)) [4, 7], а также при получении обратных мел-частотных кепстральных коэффциентов (inverted mel-frequency cepstral coefficients (IMFCC)). Для расчета IMFCC перекрывающиеся треугольные фильтры линейно размещаются по шкале инвертированного мела [20]. Это означает, что IMFCC будут описывать более высокочастотную область спектра, т.е. противоположную человеческому восприятию. В свою очередь для расчёта кепстральных коэффициентов прямоугольного набора фильтров (rectangular filter bank cepstral coefficients (RFCC)) [5] используют банк однородных неперекрывающихся прямоугольных фильтров, распределенных по линейной шкале частот.

MFCC и LFCC часто дополняются их первой и второй производными для представления временной структуры входных данных. Они называются дельта- и двойной дельта-функциями [4] соответственно. На практике они часто рассчитываются с помощью аппроксимации производной с помощью центральной разности:

$$d(t) = \frac{\sum_{n=1}^{N} n\left[c(t+n) - c(t-n)\right]}{2\sum_{n=1}^{N} n^2} \forall t = 0,\ldots,T-1, \qquad (4)$$

где $d(t)$ – дельта в момент времени t; N – определяемая пользователем длина окна для вычисления дельты; а c – либо MFCC/LFCC, либо дельта-функции (при расчете двойных дельта-функций).

При использовании банка гамматон-фильтров в процессе расчета мел-кепстральных коэффициентов получают еще один вид признаков, характеризующих голосовой сигнал, а именно гамматон-частотные кепстральные коэффициенты (gammatone frequency cepstral coefficients (GFCC)) [6]. Гамматоновый фильтр представляет собой линейный фильтр, описываемый импульсной характеристикой, которая является произведением гамма-распределения и синусоидального тона.

Кепстральные коэффициенты кохлеарного фильтра с мгновенной частотой (cochlear filter cepstral coefficients with instantaneous frequency (CFCCIF)), впервые предложенные в [16], показали лучшие результаты в составе системы автоматической верификации диктора на ASVspoof 2015 (конкурса по разработке мер противодействия спуфинговым атакам на системы автоматической верификации диктора). CFCCIF является объединением ранее используемых коэффициентов кохлеарного фильтра (cochlear filter cepstral coefficients (CFCC)), предложенных в [17] с мгновенной частотой.

Больше полезной информации, отражающей особенности человеческой речи, в сравнении с другими кепстральными коэффициентами (например, с МКК) согласно исследованию [18] возможно извлечь с помощью так называемых комплексных кепстральных коэффициентов (Complex Cepstral coefficients (CCC)). Комплексный кепстр определяется как обратное преобразование Фурье логарифма быстрого преобразования Фурье речевого сигнала. Полученные комплексные кепстральные коэффициенты (ККК) характеризуют медленно и быстро меняющиеся компоненты речи. Медленно меняющиеся компоненты (например, вклад высоты тона) концентрируются в верхней части кепстрального интервала, тогда как быстро меняющиеся компоненты (например, вклад фильтра речевого тракта) концентрируются в нижней части.

Отличным от рассмотренных выше подходов методом извлечения полезной информации из речевого сигнала является константное Q-преобразование и сопутствующие ему коэффициенты (constant Q-cepstral coefficients (CQCC)) [13]. В отличие от подхода Фурье, константное Q-преобразование дает более высокое частотное разрешение на более низких частотах и более высокое временное разрешение на более высоких частотах. Это становится возможным за счет повышения добротности. Как альтернатива используется (infinite impuls response constant-Q transform cepstrum (ICQC)) [8].

Линейное предсказание

Одним из наиболее эффективных методов, применяемых в задачах цифровой обработки речи, является линейное предсказание. Суть метода сводится к цифровой фильтрации оцифрованной речи, при которой текущий отсчет сигнала может быть «предсказан» (например, при автоматическом синтезе речи) линейной комбинацией прошлых значений выходной последовательности и настоящих, а также прошлых значений входной последовательности. Понятие «линейная комбинация» означает сумму произведений известных дискретных отсчетов сигнала (входных и выходных), умноженных на соответствующие коэффициенты линейного предсказания для предсказания (определения) неизвестного выходного отсчета. При линейном предсказании основная задача анализа речи – найти коэффициенты этой линейной комбинации, которые дают минимальную ошибку предсказания на участке анализа сигнала. Модель сигнала, наиболее часто используемая при линейном предсказании, сводится к получению неизвестного отсчета $x(n)$ без учета предыдущих входных воздействий на выходе некоторой системы

$$x(n) = \sum_{k=1} a_k x(n-k) + p, \qquad (5)$$

где p – число коэффициентов, используемых в модели; k – *коэффициенты линейного предсказания* (linear prediction coefficients (codes) (LPC)); G – коэффициент усиления, определяющий вклад в линейную комбинацию входного отсчета; $u(n)$ – текущий входной отсчет.

Задача анализа оцифрованной речи сводится к определению коэффициентов k и G этой модели.

Переход от коэффициентов линейного предсказания ккепстральным коэффициентам линейного предсказания (Linear Predictive Cepstral Coefficients (LPCC)), использующихся как значимый признак речевого сигнала, обеспечивается с помощью рекурсивного алгоритма, представленного в статье [5].

Однако более эффективным в вычислительном отношении анализ речи, основанный на линейном предсказании, выполняется при нахождении кепстральных коэффициентов перцептивного линейного предсказания (perceptual linear prediction cepstral coefficients (PLPCC)) [21]. Перцептивное линейное предсказание оперирует тремя основными понятиями психофизики слуха: 1) критическими полосами слуха, 2) кривыми равности громкости и 3) соотношением интенсивности-громкости.

Спектральный центроид

Так как при создании синтетической речи алгоритмы фокусируются на воспроизведении идентичной огибающей речевого спектра, то открытыми для исследования остаются более тонкие аспекты сигнала, располагающиеся в поддиапазонах. Такую информацию не способны отобразить рассмотренные выше кепстральные характеристики, однако с этой задачей хорошо справляются частота и амплитуда спектрального цендроида (spectral centroid frequency (SCF) и spectral centroid magnitude (SCM)) [19]. С использованием данных характеристик вычисляются соответственно коэффциенты частоты спектрального цендроида (spectral centroid frequency coefficients (SCFC)) и коэффициенты амплитуды спектрального цендроида (spectral centroid magnitude coefficients (SCMC)).

Другие значимые признаки

Одним из наиболее часто встречающихся параметров описания речевого сигнала является спектральный поток, отражающий покадровое изменение спектра мощности. Он вычисляется как евклидово расстояние между нормализованным спектром мощности последовательных кадров. Пример расчета коэффициентов спектрального потока (subband spectral flux coefficients, SSFC) при веден в [7].

Частота изменения знака в сигнале называется частотой пересечения нуля (zero crossing rate, ZCR) [1], также представляющей собой часто используемый параметр речевого сигнала. Сигналы с более низкой частотой имеют более низкую скорость пересечения нуля из-за меньшего количества колебаний в секунду по сравнению с сигналами с более высокой частотой.

$$\frac{1}{W_L}\sum_{n=1}^{W_L}\left|sgn\left[x(n)\right]-sgn\left[x(n-1)\right]\right|, \qquad (6)$$

где $x(n)$ – звуковой сигнал; WL – длина окна; а sgn – сигнум-функция (кусочно-постоянная функция действительного аргумента).

Заключение

Возможность детекции синтетических примеров речевого сигнала, в общем виде представляющих собой состязательную атаку на систему верификации диктора (дипфейк), позволяет значительно повышать уровень надежности подобных систем. Как показано в данном обзоре, одним из главных аспектов при проектировании защиты от атак подмены голоса диктора является извлечение наиболее значимых признаков из речевого сигнала. Такие признаки позволяют однозначно верифицировать диктора и несут в себе информацию, отличающую голос реального человека от синтетической подделки.

Среди рассмотренных признаков наибольшую популярность в задачах верификации диктора приобрели различные кепстральные коэффициенты. Несмотря на это, не менее эффективными в подобных задачах оказываются и классические речевые признаки (основная частота, формантные частоты), а также такие, простые на первый взгляд, параметры, как спектрограммы (мел-спектрограмы) и частота пересечения нуля.

В свою очередь, линейное предсказание на протяжении долгого времени не теряет своей актуальности в задачах цифровой обработки речи и часто используется для оценки периода основного тона, формант и других основных параметров речи [6], а наиболее перспективными с исследовательской точки зрения признаками можно считать коэффициенты, извлекаемые при работе с спектральным центроидом и спектральным потоком.

Список литературы

1. Janavi Khochare, Chaitali Joshi, Bakul Yenarkar [et al.]. A Deep Learning Framework for Audio Deepfake Detection // Arabian Journal for Science and Engineering. 2022. Vol. 47. P. 3447–3458.

2. Thanh Thi Nguyena, Quoc Viet Hung Nguyenb, Dung Tien Nguyena [et al.]. Deep Learning for Deepfakes Creation and Detection: A Survey. 2022.

3. Рахманенко И. А., Шелупанов А. А., Костюченко Е. Ю. Автоматическая верификация диктора по произвольной фразе с применением сверточных глубоких сетей доверия // Computer Optics. 2020. Vol. 44. P. 596–605.

4. Frank J., Schönherr L.WaveFake: a data set to facilitate audio deepfake detection // 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Track on Datasets and Benchmarks, 2021.

5. Balamurali B. T., Kin Wah Edward Lin, Simon Lui [et al.]. Towards robust audio spoofing detection: a detailed comparison of traditional and learned features. 2016. Vol. 6.

6. Судьенкова А. В. Обзор методов извлечения акустических признаков речи в задаче распознавания диктора // Сборник научных трудов НГТУ. 2019. № 3–4 (96). С. 139–164.

7. Sahidullah Md., Kinnunen T., Hanilci C. A comparison of features for synthetic speech detection // INTERSPEECH. 2015. P. 2087–2091.

8. Sahidullah Md., Delgado H., Todisco M. [et al.]. Introduction to Voice Presentation Attack Detection and Recent Advances // Handbook of Biometric Anti-Spoofing. 2019. P. 321–361.

9. Рахманенко И. А., Мещеряков Р. В. Анализ идентификационных признаков в речевых данных с помощью GMM-UBM системы верификации диктора // Computer Optics. 2020. Vol. 44.

10. Muckenhirn H., Magimai-Doss M., Marcel S. End-to-end convolutional neural network-based voice presentation attack detection // IEEE Int. Joint Conf. on Biometrics (IJCB). Denver, Colorado, USA, 2017. P. 335–341.

11. Yanick Lukic, Carlo Vogt, Oliver Durr, Thilo Stadelmann. Speaker identification and clustering using convolutional neural networks // IEEE International workshop on machine learning for signal processing. Salerno, Italy, 2016.

12. Sreenivas Sremath Tirumala. Speaker identification features extraction methods: A systematic review // Expert Systems With Applications 90. 2017. P. 250–271.

13. Mardu R. Kamle. Advances in anti-spoofing: from the perspective of ASVspoof challenges // SIP. 2020. Vol. 9, e2. P. 1–18.

14. Hong Yu., Tan Z.-H., Ma Z. [et al.]. Spoofing detection in automatic speaker verification systems using DNN classifiers and dynamic acoustic features // IEEE Trans. NeuralNetw. Learn. Syst. 2017. Vol. 29 (10). P. 4633–4644.

15. Honglak Lee, Peter Pham, Yan Largman, and Andrew YNg. Unsupervised feature learning for audio classification using convolutional deep belief networks // Advancesin neural information processing systems. 2009. P. 1096–1104.

16. Tanvina B. P., Hemant A. P. Combining Evidences from Mel Cepstral, Cochlear Filter Cepstral and Instantaneous Frequency Features for Detection of Natural vs. Spoofed Speech // INTERSPEECH. Dresden, Germany. 2015. P. 2062–2066.

17. Li Q. An auditory-based transform for audio signal processing // 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. 2009. P. 181–184.

18. Balamurali B. T. Nair, Esam A. S. Alzqhoul, Bernard J. Guillemin. Comparison between Mel-Frequency and Complex Cepstral Coefficients for Forensic Voice Comparison using a Likelihood Ration Framework // Proceedings of the World Congress on Engineering and Computer Science. Vol. I WCECS 2014, San Francisco, USA, 2014.

19. Jia Min Karen Kua, Tharmarajah Thiruvaran, Mohaddeseh Nosratighods [et al.]. Investigation of Spectral Centroid Magnitude and Frequency for Speaker Recognition // The Speaker and Language Recognition Workshop 28 June – 1 July 2010). Brno, Czech Republic, 2010.

20. Sharma D., Ali I. A modified MFCC feature extraction technique for robust speaker recognition // Proceedings of International Conference on Advances in Computing, Communications and Informatics (ICACCI 2015). Kochi, India, 2015. P. 1052–1057.

21. Hermansky H. Perceptual linear predictive (PLP) analysis of speech // The Journal of the Acoustical Society of America. 1990. Vol. 87, № 4. P. 1738–1752.

Для цитирования: Панфилова И. Е., Сулавко А. Е. Обзор признаков, извлекаемых из речевых сигналов с целью распознавания состязательных примеров // Безопасность информационных технологий : сб. науч. ст. по материалам IV Всерос. науч.-техн. конф. : в 2 т. Пенза : Изд-во ПГУ, 2022. Т. 1. С. 67–77.

УДК 519.24

ИСПОЛЬЗОВАНИЕ ЭФФЕКТА НОРМАЛИЗАЦИИ ЗАКОНА РАСПРЕДЕЛЕНИЯ ИНТЕГРОДИФФЕРЕНЦИАЛЬНОГО СТАТИСТИЧЕСКОГО КРИТЕРИЯ КРАМЕРА – фон МИЗЕСА

А. П. Иванов[1], Е. А. Малыгина[2], К. А. Перфилов[3], С. Е. Вятчанин[4]

[1,2,3,4] Пензенский государственный университет, г. Пенза

Аннотация. Рассматривается проблема статистического анализа малых выборок с использованием интегродифференциального критерия Крамера – фон Мизеса. Для двух интервалов построено почти аналитическое описание значений *dKfM*-критерия проверки статистических гипотез, что создает условия для его практического применения при статистической оценке малых биометрических выборок.

Ключевые слова: статистический анализ малых выборок, проверка гипотезы нормальности, интегродифференциальный критерий Крамера – фон Мизеса

USING THE EFFECT OF NORMALIZATION OF THE LAW OF DISTRIBUTION OF THE INTEGRO-DIFFERENTIAL STATISTICAL CRITERION «KRAMER-VON MISES»

A. P. Ivanov[1], E. A. Malygina[2], K. A. Perfilov[3], S. E. Vyatchanin[4]

[1,2,3,4] Penza State University, Penza

Abstract. The paper deals with the problem of statistical analysis of small samples using the Cramer-von Mises integro-differential test. For two intervals, an almost analytical description of the values of the dKfM criterion for testing statistical hypotheses is constructed, which creates conditions for its practical application in the statistical evaluation of small biometric samples.

Keywords: statistical analysis of small samples, testing of the hypothesis of normality, Cramer-von Mises integro-differential test

Известно, что в соответствие с основной статистической теоремы, суммирование независимых данных приводит к их нормализации [1−3]. Результатами проведенных численных экспериментов подтверждено, что статистический критерий Крамера – фон Мизеса обладает более выраженным эффектом нормализации данных, чем критерий хи-квадрат [4−6]. Еще больше эффект нормализации работает для интегро-дифференциального критерия Крамера – фон Мизеса [7]. Эта ситуация иллюстрируется рис. 1.

© Иванов А. П., Малыгина Е. А., Перфилов К. А., Вятчанин С. Е., 2022

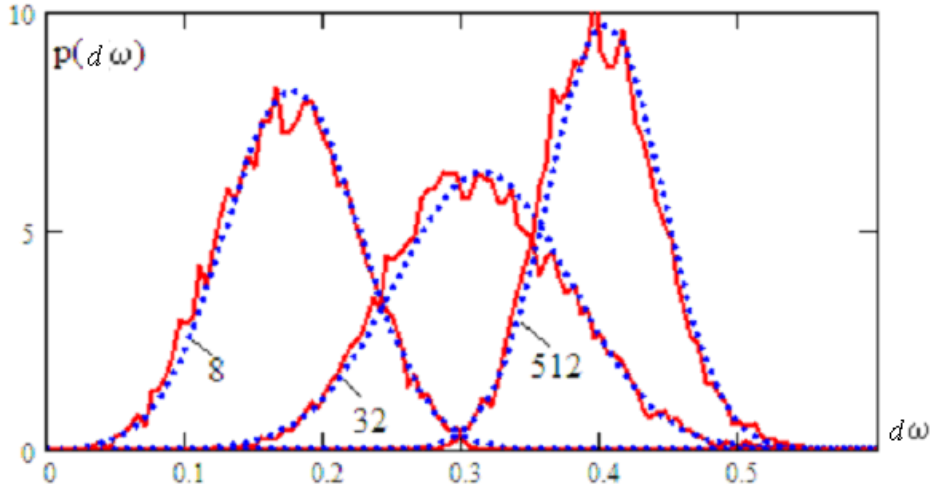

Рис. 1. График, демонстрирующий близость к нормальному закону распределения (пунктир) значений интегро-дифференциального критерия (непрерывные линии) для выборки из 8, 32, 512 опытов

Из данных рис. 1 видно, что из-за эффекта нормализации данных можно строить аналитическое описание плотности распределения значений как нормальное распределение, математическое ожидание и стандартное отклонение которого являются функциями размеров обучающей выборки:

$$p(d\omega, N) = \frac{1}{\sigma(N)\cdot\sqrt{2\cdot\pi}}\cdot\exp\left\{\frac{-\left(E(N)-d\omega\right)^2}{2\cdot\left(\sigma(N)\right)^2}\right\}. \qquad (1)$$

Для того, чтобы получить аналитическое описание плотности распределения значений (1), необходимо найти аналитическое описание функции $E(N)$ и функции $\sigma(N)$. Отметим, что функция математического ожидания – $E(N)$ для значений *dKfM*-критерия является монотонно возрастающей, что упрощает задачу. Иная ситуация возникает для функции стандартного отклонения – $\sigma(N)$. Данная функция монотонно возрастает для малых выборок (примерно до 20 примеров) (рис. 2).

Из данных рис. 2 видно, что для выборок в 21 пример и более функция $\sigma(N)$ монотонно падает. В связи с этим поставленную задачу будем решать в два этапа. На первом этапе построим аналитическое описание для выборок от 8 до 20 примеров, а на втором – от 21 примера и более.

При этом функция стандартного отклонения для малых выборок всегда монотонно увеличивается (рис. 3).

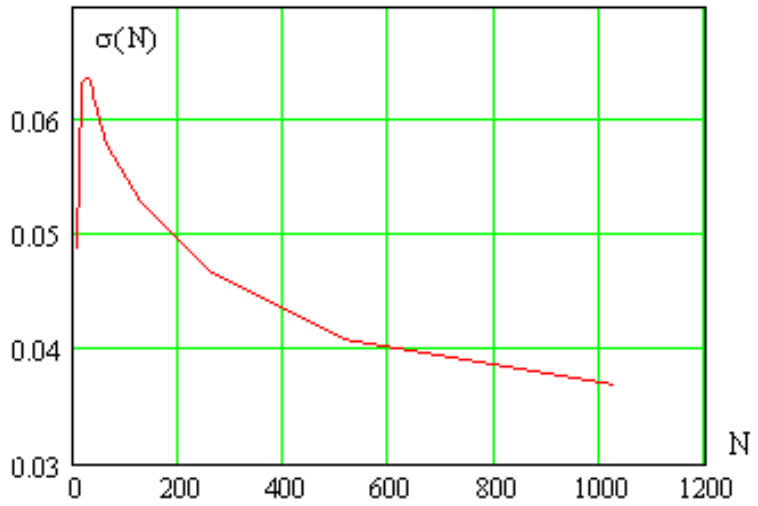

Рис. 2. График функции стандартного отклонения
для нормальной закона распределения значений критерия *dKfM*

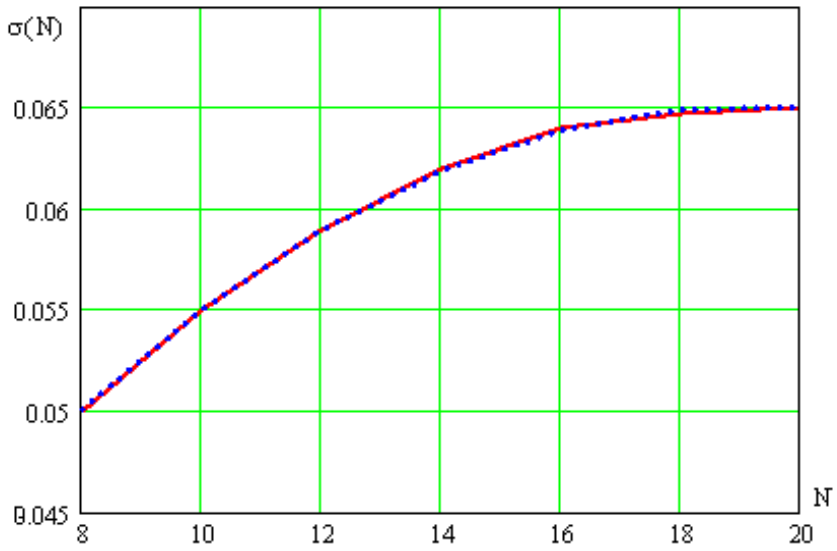

Рис. 3. График монотонно увеличения стандартного отклонения
в интервале от 8 до 20 опытов

Такая функция хорошо описывается полиномом второго порядка (пунктирная линия):

$$\sigma(N) = 0,065 - 0,0002 \cdot (N-20)^1 - 0,0001 \cdot (N-20)^2. \quad (2)$$

Функция математического ожидания так же является монотонно увеличивающейся и ее приближение полиномом второго порядка:

$$E(N) = 0,28 - 0,0005 \cdot (N-20)^1 - 0,0007 \cdot (N-20)^2. \quad (3)$$

Функция монотонного снижения значения функции стандартного отклонения значений *dKfM*-критерия хорошо приближается в логарифмическом пространстве объема выборки:

$$\sigma(N) = 0,088 - 0,0051 \cdot \log_2(N). \quad (4)$$

Приближение функции $\sigma(N)$ в логарифмическом пространстве связано с медленным градиентом ее уменьшения. Приближение функции математического ожидания в этом же интервале описывается гораздо более крутой функцией, имеющей тенденцию к «насыщению» со слабым линейным ростом. Для ее описания приходится использовать полином второго порядка от объема выборки в логарифмическом пространстве:

$$E(N) = 0,019 + 0,079 \cdot \log_2(N) - 0,004 \cdot \{\log_2(N)\}^2. \quad (5)$$

Дифференциальный вариант критерия хорошо описывается одним нормальным законом. Таким образом, возможно построение для двух интервалов почти аналитическое описание значений *dKfM*-критерия проверки статистических гипотез, что создает условия для его практического применения при статистической оценке малых биометрических выборок.

Список литературы

1. Р 50.1.037–2002. Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Ч. I. Критерии типа χ2. М. : Госстандарт России, 2001. 140 с.

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

3. Р 50.1.037–2002. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Ч. II. Непараметрические критерии. М. : Госстандарт России, 2002. 123 с.

4. Иванов А. И., Газин А. И., Вятчанин С. Е., Перфилов К. А. Сравнение мощности хи-квадрат критерия и критерия Крамера – фон Мизеса для малых тестовых выборок биометрических данных // Надежность и качество сложных систем. 2016. № 2 (14). С. 67–72.

5. Вятчанин С. Е., Иванов А. И., Газин А. И., Перфилов К. А. Подавление шумов квантования биометрических данных при использовании многомерного критерия Крамера – фон Мизеса // Проблемы информационной безопасности. Компьютерные системы. 2016. № 2. С. 21–28.

6. Вятчанин С. Е., Иванов А. И., Малыгина Е. А., Перфилов К. А. Сравнение мощности критерия среднего геометрического и Крамера –

фон Мизеса на малых выборках биометрических данных // Модели, системы, сети в экономике, технике, природе и обществе. 2016. № 2 (18). С. 155–163.

7. Иванов А. И., Малыгина Е. А., Серикова Ю. И. [и др.]. Обоснование и выбор статистических критериев для корректной оценки данных малых выборок биометрических образов // Труды Междунар. симп. Надежность и качество. 2018. Т. 2. С. 456–459.

Для цитирования: Иванов А. П., Малыгина Е. А., Перфилов К. А., Вятчанин С. Е. Использование эффекта нормализации закона распределения интегро-дифференциального статистического критерия Крамера – фон Мизеса» // Безопасность информационных технологий : сб. науч. ст. по материалам IV Всерос. науч.-техн. конф. : в 2 т. Пенза : Изд-во ПГУ, 2022. Т. 1. С. 78–82.

УДК 004.056;004.89;57.087

РАЗРАБОТКА УЧЕБНОГО АППАРАТНО-ПРОГРАММНОГО КОМПЛЕКСА ДЛЯ ИССЛЕДОВАНИЯ ПОДАВЛЕНИЯ СИГНАЛОВ БЕСПРОВОДНОЙ СВЯЗИ

К. О. Кильдюшкин[1], Д. Е. Баринов[2], А. П. Иванов[3]

[1,2,3] Пензенский государственный университет, г. Пенза

Аннотация. Разработана структура учебного аппаратно-программного комплекса для исследования подавления беспроводной передачи данных. Проведено тестирование работоспособности разработанного комплекса. Экспериментально получена зависимость вероятности приема пакетов от различных параметров.

Ключевые слова: беспроводная передача данных, вероятность правильного приема пакетов, программно-аппаратный комплекс

DEVELOPMENT OF EDUCATIONAL HARDWARE AND SOFTWARE COMPLEX FOR WIRELESS SIGNAL SUPPRESSION STUDIES

K. O. Kildyushkin[1], D. E. Barinov[2], A. P. Ivanov[3]

[1,2,3] Penza State University, Penza

Abstract. The structure of an educational hardware-software complex for studying the suppression of wireless data transmission has been developed. The performance testing of the developed complex was carried out. The dependence of the packet reception probability on various parameters has been experimentally obtained.

Keywords: wireless data transmission, probability of correct packet reception, software and hardware complex

В современном мире, ввиду большого количества устройств, проводные соединения не позволяют дать такое удобство пользователю, как беспроводные. Любой смартфон имеет в своем функционале такие варианты беспроводной связи, как Wi-Fi и Bluetooth. С их помощью можно получить доступ в сеть Internet, передавать данные с одного устройства на другое, связываться с другими пользователями. Данные стандарты работают в нелицензируемом частотном диапазоне беспроводной связи ISM 2,4 ГГц. Преимуществами данного диапазона над проводными являются:

© Кильдюшкин К. О., Баринов Д. Е., Иванов А. П., 2022

– независимость от препятствий окружающей среды, таких, как водные преграды, автомобильные и железные дороги и т.д.;

– стоимость комплексов приемо-передачи значительно ниже в связи с отсутствием дорогостоящих проводных линий связи.

– удобство пользователя, в связи с отсутствием многочисленного количества проводов.

Но данный вид связи не обходится и без недостатков:

– зависимость от препятствий на пути распространения радиосигналов (стены, окна, двери), количества устройств, работающих на тех же частотах, значительно влияют на качество приема данных;

– из первого недостатка следует, что реальная скорость передачи намного ниже, чем при проводной линии связи, причиной этому является большой уровень помех;

– перехват информации может осуществляться на значительном расстоянии от линии передачи (в случае проводной линии связи, злоумышленнику необходимо находиться в непосредственной близости от линии передачи).

Наиболее значительным из всех представленных недостатков является малая помехозащищенность.

Основываясь на вышеперечисленном, была сформулирована цель проекта – создание учебного аппаратно-программного комплекса для исследования подавления сигналов беспроводной связи. Данный комплекс разрабатывается для изучения студентами влияния помех на качество передачи информации по радиоканалам, а также рассмотрении вариантов борьбы с помехами [1].

Разработанный комплекс аппаратно включает в себя три контроллера (два Arduino Mega 2560 и один Arduino Nano 3.0), три радиомодуля (два NRF24L01 и один NRF24L01+PA+LNA), дисплейный модуль и блока кнопок. Данные компоненты были выбраны в связи с низкой стоимостью и доступностью, а также из-за достаточных характеристик для комплекса. Также, для работы комплекса, необходим хотя бы один компьютер (АРМ), имеющий два свободных USB порта для подключения приемника и передатчика.

Данные компоненты входят в три части комплекса – передатчик, приемник и генератор помех. Структура комплекса показана на рис. 1.

Аппаратно-программный комплекс имеет следующие параметры передатчика/приемника:

– мощность передатчика – –18, –12, –6 и 0 дБм;

– частота передачи – от 2400 до 2524 МГц;

– скорость передачи – 250, 1024 и 2048 кбит/с;

– размер передаваемого пакета – от 1 до 32 байт;

– количество передаваемых пакетов – 100, 500, 1000, 5000 и 10000 пакетов;

– режим обратной связи – включена или выключена;

– количество попыток повторной отправки пакетов (настраивается при включенном режиме обратной связи) – от 0 до 10.

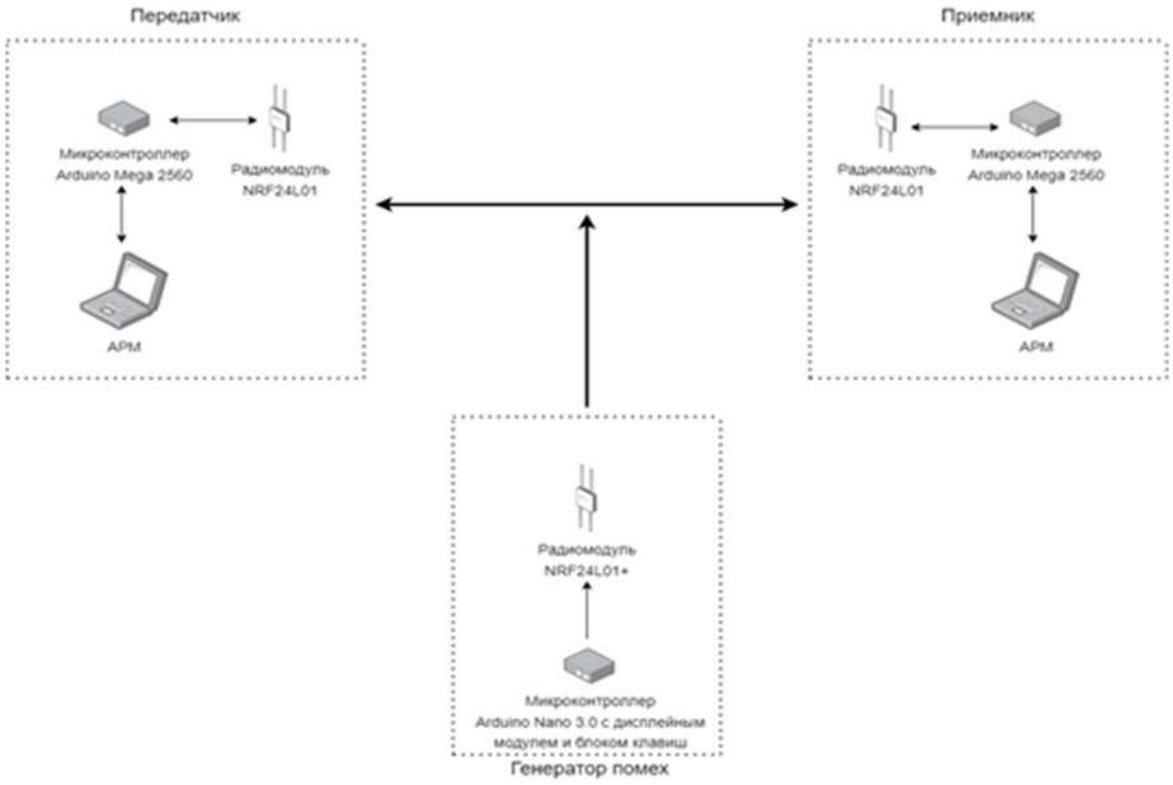

Рис. 1. Структура аппаратно-программного комплекса

В генераторе помех реализована возможность установить следующие параметры:

– частота передачи – от 2400 до 2524 МГц;

– режим работы – постоянная частота, последовательное изменение частоты, случайное изменение частоты (в последних двух вариантах также устанавливается диапазон изменения частот и время работы на определенной частоте);

– мощность передачи – –18, –12, –6 и 0 дБм;

– скорость передачи – 250, 1024 и 2048 кбит/с.

В качестве защиты от помех используется повторная передача информационного пакета, при отсутствии квитанции от приемника.

Данный набор параметров обеспечивает большое количество вариантов исследований подавления сигналов беспроводной связи.

Также стоит отметить, что важной характеристикой для исследования будет являться расстояние между приемником, передатчиком и подавителем, которые будут существенно влиять на результаты исследований.

Для управления передатчиком/приемником был разработан графический интерфейс, устанавливаемый на АРМ, с помощью которого можно управлять настройками комплекса. Меню настройки параметров графического интерфейса показано на рис. 2.

Для управления настройками подавителя используется дисплей с блоком кнопок, что делает его независимым от расположения АРМ.

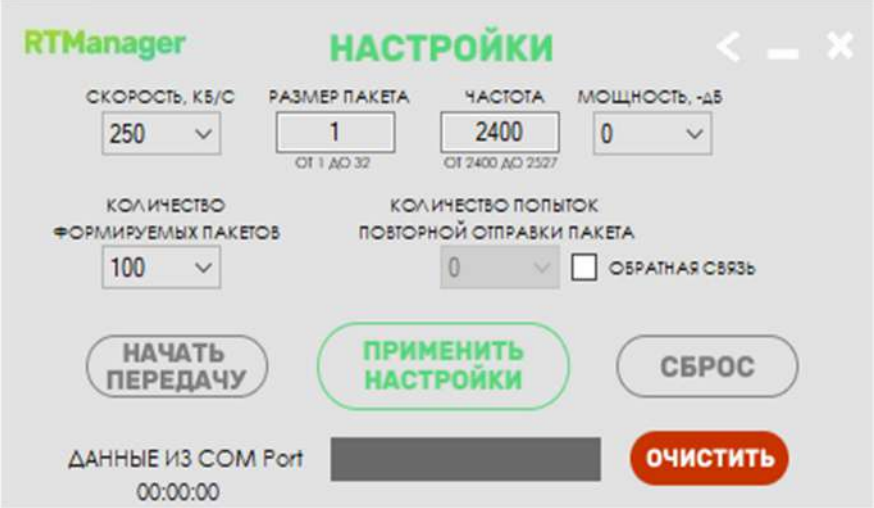

Рис. 2. Меню настройки параметров графического интерфейса

С помощью разработанного аппапатно-программного комплекса были проведены следующие исследования подавления сигналов беспроводной связи:

– исследование влияния отношения сигнал/шум и скорости передачи пакетов на вероятность успешного приема пакетов;

– исследование влияния влияния количества попыток повторной отправки пакетов и скорости передачи на вероятность успешного приема пакетов;

– исследование влияния размера пакета и скорости передачи на вероятность успешного приема пакета.

Результат исследования влияния отношения сигнал/шум и скорости передачи пакетов на вероятность успешного приема пакетов приведен на рис. 3.

График показывает, что разработанный аппаратно-программный комплекс обеспечивает передачу с вероятностью успешного приема пакета 0,8:

– на скорости 250 кбит/с при отношении сигнал/шум более 14,1 дБм;

– на скорости 1024 кбит/с при отношении сигнал/шум более 18,3 дБм;

– на скорости 2048 кбит/с при отношении сигнал/шум более 20 дБм.

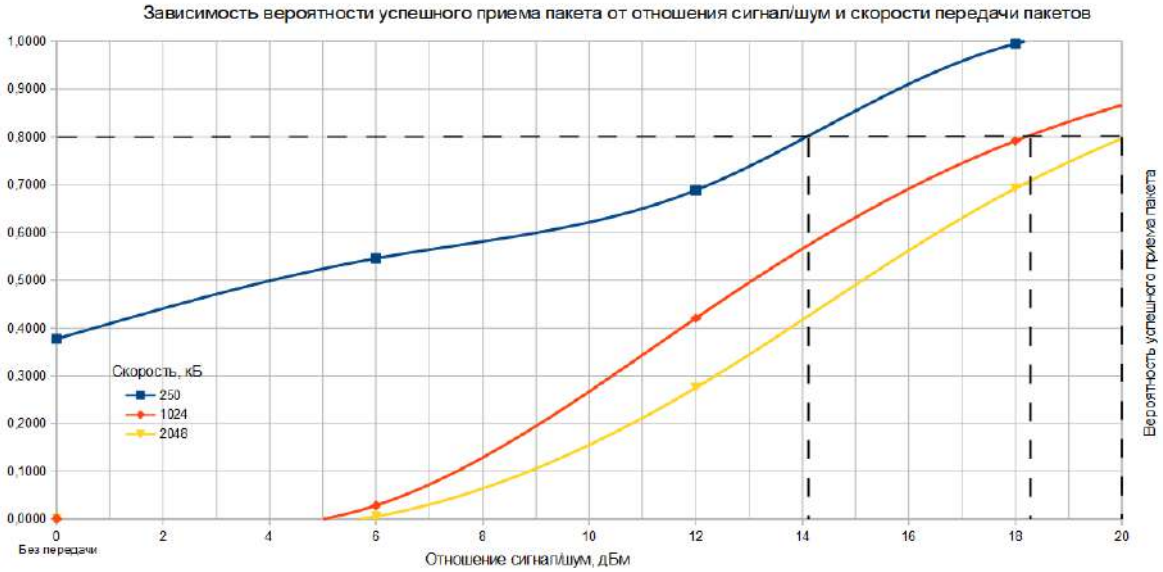

Рис. 3. График зависимости вероятности успешного приема пакетов от отношения сигнал/шум и скорости передачи пакетов

Далее было проведено исследование влияния скорости передачи и количества пакетов обратной связи на вероятность успешного приема пакетов, результат которого показан на рис. 4.

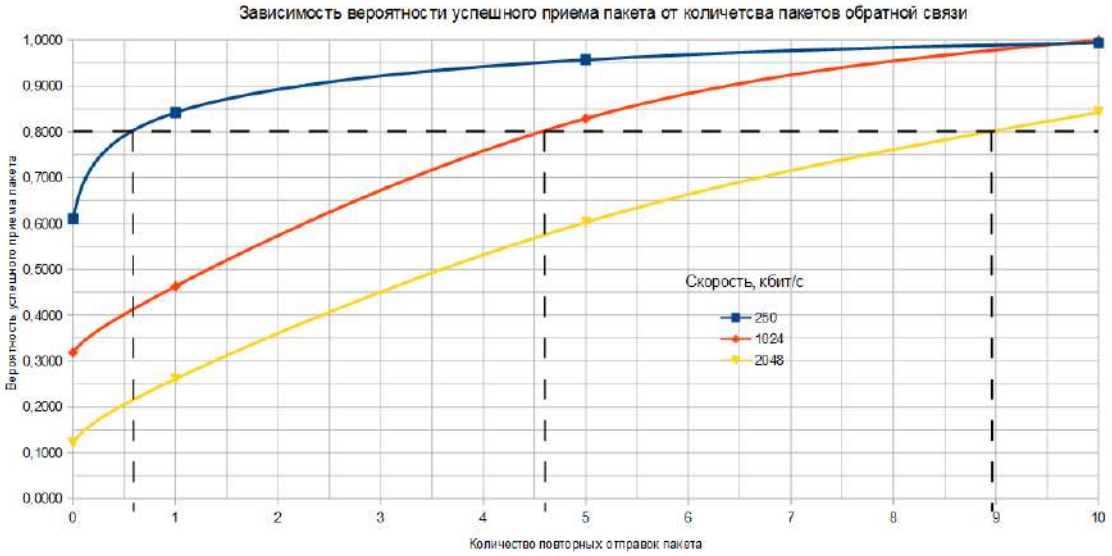

Рис. 4. График зависимости вероятности успешного приема пакета от количества пакетов обратной связи и скорости

Построенный график, показанный на рис. 4, показывают, что увеличение количества повторных отправок пакета приводит к увеличению вероятности успешного приема пакета. Но стоит учитывать, что увеличение скорости передачи отрицательно влияет на вероятность успешного приема. Разработанный комплекс позволяет вести передачу с вероятность успешного приема 0,8 при:

– скорости передачи 250 кбит/с и не менее 1 повторной отправок пакета;

– скорости передачи 1024 кбит/с и не менее 5 повторных отправок пакета;

– скорости передачи 2048 кбит/с и не менее 9 повторных отправок пакета.

Заключительным стало проведение исследования влияния изменения скорости передачи пакетов и размера пакета на вероятность успешного приема пакетов, результаты которого показаны на рис. 5.

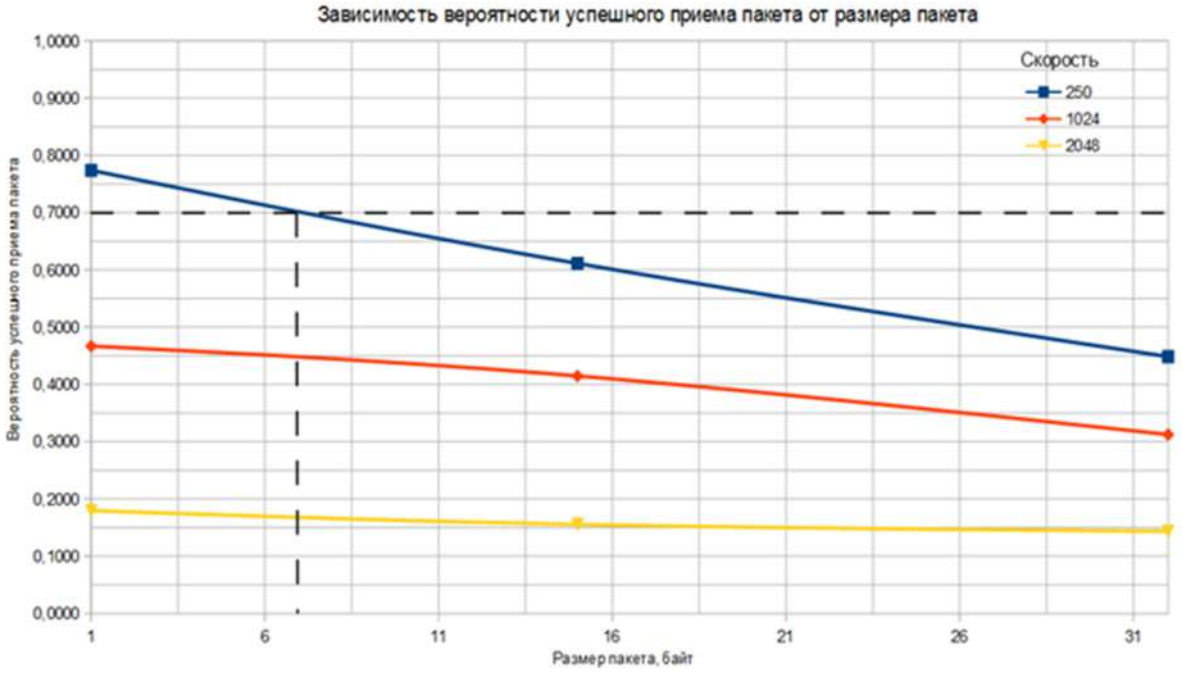

Рис. 5. График зависимости скорости передачи
и размера пакета на вероятность успешного приема

На графике наблюдается уменьшение вероятности успешного приема при увеличении размера пакета. Аппаратно-программный комплекс обеспечивает передачу пакетов с вероятностью успешного приема 0,7 только на скорости 250 кбит/с при размере пакета не более 7 байт.

Проведенные исследования не являются всеми возможными, которые позволяет провести аппаратно-программный комплекс.

Таким образом для студентов, выполняющими лабораторный практикум, появляется возможность самим разрабатывать варианты проведения исследований.

С использованием разработанного учебного аппаратно-программного комплекса студенты смогут получить следующие компетенции [2]:

– способность формулировать задачи, планировать и проводить исследования, в том числе эксперименты и математическое моделирование объектов, явлений и процессов телекоммуникационных систем, включая обработку и оценку достоверности их результатов (ОПК-12);

– способность оценивать технические возможности, анализировать угрозы и вырабатывать рекомендации по построению элементов информационно-телекоммуникационной инфраструктуры с учетом обеспечения требования информационной безопасности (ОПК-13).

Список литературы

1. Платонов В. Д., Киселев В. Ю., Иванов А. П. Разработка учебного аппаратно-программного комплекса беспроводной передачи данных // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. (г. Пенза, 3 июня 2020 г.). Пенза : Изд-во ПГУ, 2020. С. 193–197.

2. Об утверждении федерального государственного стандарта высшего образования – специалитет по специальности 10.05.02 Информационная безопасность телекоммуникационных систем : приказ Министерства науки и высшего образования Российской Федерации № 1458 от 26.11.2020. URL: https://fgosvo.ru/uploadfiles/FGOS%20VO%203++/Spec/100502_C_3_15022021.pdf (дата обращения: 04.05.2022).

Для цитирования: Кильдюшкин К. О., Баринов Д. Е., Иванов А. П. Разработка учебного аппаратно-программного комплекса для исследования подавления сигналов беспроводной связи // Безопасность информационных технологий : сб. науч. ст. по материалам IV Всерос. науч.-техн. конф. : в 2 т. Пенза : Изд-во ПГУ, 2022. Т. 1. С. 83–89.

УДК 519.24;53;57.017

КОМПАКТНЫЕ ИСПОЛНЯЕМЫЕ ПРИЛОЖЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ SIM-КАРТ: ПРОГНОЗ ЭКОНОМИИ ЧИСЛА БИНАРНЫХ НЕЙРОНОВ ПРИ ИХ ЗАМЕНЕ БОЛЕЕ СЛОЖНЫМИ Q-АРНЫМИ ИСКУССТВЕННЫМИ НЕЙРОНАМИ

А. И. Иванов[1], А. П. Иванов[2], В. А. Цимбал[3]

[1]Пензенский научно-исследовательский электротехнический институт, г. Пенза
[2]Пензенский государственный университет, г. Пенза
[3]Филиал Военной академии Ракетных войск стратегического назначения имени Петра Великого, г. Серпухов Московской области

Аннотация. При разработке исполняемых приложений для SIM-карт приходится ориентироваться на малые вычислительные ресурсы их контроллеров. Показано, что следующее поколение сетей искусственных нейронов может давать значительную экономию памяти, потребляемой энергии, времени выполнения приложений, если заменить в однослойной сети обычные бинарные нейроны на их q-арные аналоги.

Ключевые слова: бинарные искусственные нейроны, экономия вычислительных ресурсов, q-арные нейроны, сокращение числа нейронов

COMPACT EXECUTABLE ARTIFICIAL INTELLIGENCE APPLICATIONS FOR SIM-CARDS: PREDICTION OF SAVINGS IN THE NUMBER OF BINARY NEURONS IN THEIR REPLACEMENT WITH MORE COMPLEX Q-ARY ARTIFICIAL NEURONS

A. I. Ivanov[1], A. P. Ivanov[2], V. A. Tsymbal[3]

[1]Penza Research Electrotechnical Institute, Penza
[2]Penza State University, Penza
[3]Branch of Military Academy of Strategic Rocket Troops after Peter the Great, Serpukhov, Moscow region

Abstract. When developing executable applications for SIM cards, one has to focus on the small computing resources of their controllers. It is shown that the next generation of networks of artificial neurons can provide significant savings in memory, energy consumption, and application execution time if ordinary binary neurons are replaced in a single-layer network with their q-ary counterparts.

© Иванов А. И., Иванов А. П., Цимбал В. А., 2022

Keywords: binary artificial neurons, saving computational resources, q-ary neurons, reduction in the number of neurons

Введение

Следует отметить, что в соответствии с Указом президента В. В. Путина № 480 от 10.10.2019 к 2030 г. Россия будет занимать одно из ведущих положений по использованию приложений искусственного интеллекта (ИИ). При этом важную роль играет низкая стоимость приложений ИИ, их безопасность и массовость. Одним из способов решения этой задач является использование нейросетевых приложений ИИ, ориентированных на вычисления в доверенной среде контроллеров SIM-карт.

Например, это могут быть приложения нейросетевой аутентификации пользователей средств мобильной связи, выполненные по действующему в РФ стандарту ГОСТ Р 52533.5 [1]. Вполне возможно, что в ближайшее время в России будет принят и введен в действие второй национальный стандарт [2] биометрико-нейросетевой аутентификации пользователей.

Для нас принципиально важным является то, что первый национальный стандарт России [1] фактически регламентирует работу с сетями простейших бинарных нейронов (персептронов). Второй стандарт [2], уже ориентирован на более сложные и, соответственно, потенциально более эффективные квадратичные нейроны Байеса с многоуровневым выходным квантователем.

Очевидно, что рост сложности обработки данных должен приводить к росту разделимости, накопленных (обогащенных), выходных данных. Как результат, для более качественных выходных данных могут быть использованы более сложные многоуровневые квантователи [3]. Именно по этой причине простейщие персептроны, выполняющие обогащение входных данных в линейном пространстве [1] оказались ориентированы на использование бинарных квантователей. Более сложные квадратичные найроны Байеса (осуществляющие более эффективное обогащение данных) при том же числе входов у каждого нейрона уже способны иметь четырех уровневые квантователи.

Очевидно, что искусственные нейронные сети и далее будут развиваться через повышение эффективности работы каждого нейрона и повышение эффективности работы нескольких последовательно расположенных нейронов в многослойных нейросетях. То есть в ближайшем будущем задача перехода от бинарных нейронов к более сложным q-арным нейронам будет становиться все более и более актуальной.

Оценка роста качества принимаемых решений при замене пяти классических статистических критериев двоичными, троичными и пятеричными искусственными нейронами

Для практики статистического анализа реальных данных (биометрии, биологии, ботаники, медицины, экономики, экологии, психологии) характерна проблема обработки малых выборок реальных данных. Так нейросетевые преобразователи биометрических данных конкретного человека в длинный код приходится обучать на 16 примерах образа «Свой». К сожалению, классические статистические критерии дают большие ошибки на малых выборках [4]. Более того, даже младшие статистические моменты (математические ожидания, стандартные отклонения, коэффициенты корреляции) на малых выборках удается вычислять только с большой погрешностью.

Одним из путей повышения достоверности принимаемых решений, например, при решении задачи проверки гипотезы нормальности является использование не одного, а нескольких статистических критериев [5]. Для определенности рассмотрим пять следующих классических статистических критериев:

- Андерсона-Дарлинга (1952 г.);
- нормированного размаха (1954 г.);
- Васичека (1976 г.);
- Фроцини (1978 г.);
- четвертого статистического момента (1984 г.).

Все перечисленные выше классические статистические критерии работают по-разному, имеют разные вероятности ошибок первого и второго рода, а также имеют разные коэффициенты корреляции между их решениями. Последнее затрудняет численное моделирование их взаимодействия. Для упрощения ситуации целесообразно выполнить задачу симметризации взаимодействия, рассматриваемых статистических критериев [6].

Симметризация задачи дает эквивалентные симметричные нейроны, которые имеют одинаковые вероятности ошибок первого и второго рода на уровне 0,251. Оценка выполнена через вычисление среднего геометрического всех подобных показателей по каждому из, рассматриваемых критериев. Усреднение модулей коэффициентов корреляции, рассматриваемых критериев дает значение 0,431 коэффициента одинаковой корреляционной сцепленности данных. В результате, мы имеем для каждого из пяти симметризованных нейронов статистическую модель представленную на рис. 1.

Рассматриваемые эквивалентные нейроны обогащают данные в разных нелинейных пространствах. Вид пространства, в котором

происходит обогащение зависит от конструкции эмулируемого статистического критерия, однако все обогатители являются достаточно «хорошими» нормализаторами. То есть, при росте числа входов у искусственных нейронов нормализуется отклики их входных обогащающих конструкций. В связи с этим, на рис. 1 отклики симметричных нейронов представлены нормальными распределениями, хотя отклики реальных нейронов для того или иного статистического критерия могут быть существенно асимметричны [5]. Замена асимметричных распределения симметричными нормальными распределениями, является еще одним упрощением задачи.

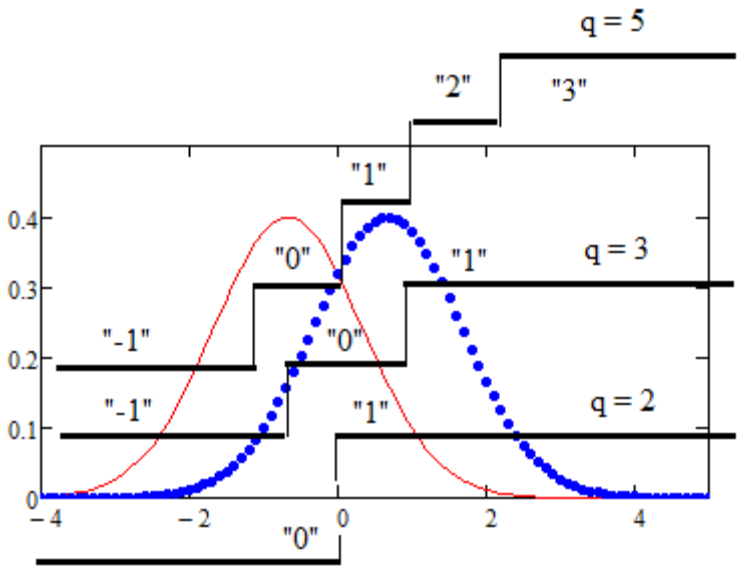

Рис. 1. Симметричная статистическая модель откликов одного из пяти нейронов на входные нормальные данные (непрерывное распределение) и равномерные входные данные (точечная линия) для малых выборок объемом в 16 опытов

Прогнозирование ожидаемых значений вероятности появления не поддающихся корректировке ошибок с ростом числа искусственных нейронов

Очевидно, что синтезировать новые статистические критерии проверки гипотезы нормальности или гипотезы равномерности [7] достаточно сложно. Куда проще повышать число выходных состояний квантователей искусственных нейронов. На рис. 1 на два разделяемых статистических распределения наложены три квантователя. Бинарный квантователь q = 2, троичный квантователь q = 3 и пятеричный квантователь q = 5.

В зависимости от вида квантователя нейросеть будет давать выходные коды в разных системах счисления. В простейшем случае

применения бинарных квантователей мы будем иметь бинарный выходной код, состоящий из 5 разрядов. Его состояние «00000» будет соответствовать ситуации, когда все 5 статистических критериев приняли решение об обнаружении нормального закона распределения.

Очевидно, что не всегда все 5 классических статистических критерия будут работать синхронно. В большинстве случаев критерии будут работать по-разному. Необходимо устранять противоречия в решении 5 разных критериев (5 разных экспертов). Проще всего это можно выполнить, подсчитывая в кодах число состояний «0». Если число состояний «0» больше числа состояний «1», то принимается решение об обнаружении нормального распределения исследуемой малой выборки в 16 опытов.

Для троичных кодов q = 3 в идеальном случае мы будем иметь код из 15 состояний «0». Для пятеричных кодов q = 5 в идеальном случае мы будем иметь код из 25 состояний «0». Для всех систем счисления свертывание избыточных кодовых конструкций выполняется по одному и тому же правилу.

Для прогнозирования достижимых вероятностей ошибок, в зависимости от числа нейронов достаточно двух точек для трех типов [8], рассмотренных в данной статье искусственных нейронов. Прогнозирование удобно выполнять в логарифмическом масштабе по двум координатами. В простейшем случае может быть использована линейная экстраполяция, как это показано на рис. 2.

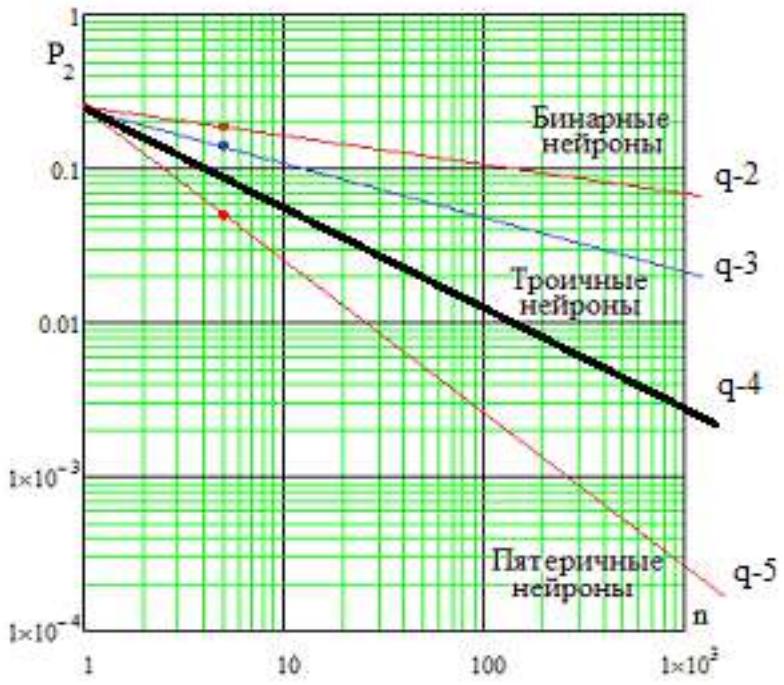

Рис. 2. Прогноз достижимых вероятностей ошибок второго рода при разном числе нейронов, имеющих многоуровневые квантователи

Из рис. 2 мы видим, что простейшие бинарные нейроны являются самыми слабыми. Даже нейросеть, состоящая из 1000 бинарных нейронов способна обеспечить решения с ожидаемой доверительной вероятностью 0,93. Троичные нейроны гораздо сильнее. Тысяча троичных нейронов должна давать решения с доверительной вероятностью 0,98, что уже приемлемо для ряда практических применений.

В случае использования нейросетевых преобразователей биометрии в код длинной в 256 бит, желательно принимать решения с доверительной вероятностью 0,997. Как видно из прогноза рис. 2 такую доверительную вероятность могут обеспечить уже 40 5-арных искусственных нейронов. Моделирование для q = 4 не выполнялось, толстая линия рис. 2 является средней между двумя соседями.

Заключение

Естественно, что выполненные нами приближения (симметризация искусственных нейронов, линейная экстраполяция в координатах двойного логарифмирования) могут приводить к существенным ошибкам предсказания. Однако авторы этой статьи убеждены, что переход от простейших бинарных нейронов к более сложным q-арным нейронам дает значительные технические преимущества.

Список литературы

1. ГОСТ Р ИСО/МЭК 19794-2–2013. Автоматическая идентификация. Идентификация биометрическая. Форматы обмена биометрическими данными. Ч. 2. Данные изображения отпечатка пальца – контрольные точки.

2. ТК 164, публичное обсуждение первой редакции стандарта ГОСТ Р...хх «Искусственный интеллект. Нейросетевые алгоритмы в защищенном исполнении. Автоматическое обучение нейросетевых моделей на малых выборках в задачах классификации».

3. Малыгина Е. А. Биометрико-нейросетевая аутентификация: перспективы применения сетей квадратичных нейронов с многоуровневым квантованием биометрических данных : препринт. Пенза : Изд-во ПГУ, 2020. 114 с.

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

5. Иванов А. П., Иванов А. И., Малыгин А. Ю. [и др.]. Альбом из девяти классических статистических критериев для проверки гипотезы нормального или равномерного распределения данных малых выборок // Надежность и качество сложных систем. 2022. № 1. С. 20–29. doi:10.21685/2307-4205-2022-1-3

6. Иванов А. И., Банных А. Г., Серикова Ю. И. Учет влияния корреляционных связей через их усреднение по модулю при нейросетевом обобщении статистических критериев для малых выборок // Надежность. 2020. № 20 (2). С. 28–34. URL: https://doi.org/10.21683/1729-2646-2020-20-2-28-34

7. Иванов А. П., Иванов А. И., Безяев А. В. [и др.]. Обзор новых статистических критериев проверки гипотезы нормальности и равномерности распределения данных малых выборок // Надежность и качество сложных систем. 2022. № 2. С. 33–44.

8. Иванов А. И., Юнин А. П., Иванов А. П. [и др.]. Мультикритериальная проверка гипотезы нормальности и равномерности малых выборок с использованием троичных и двоичных искусственных нейронов // Надежность и качество сложных систем. 2022. № 3. С. 70–77. doi:10.21685/2307-4205-2022-3-9

Для цитирования: Иванов А. И., Иванов А. П., Цимбал В. А. Компактные исполняемые приложения искусственного интеллекта для sim-карт: прогноз экономии числа бинарных нейронов при их замене более сложными q-арными искусственными нейронами // Безопасность информационных технологий : сб. науч. ст. по материалам IV Всерос. науч.-техн. конф. : в 2 т. Пенза : Изд-во ПГУ, 2022. Т. 1. С. 90–96.

УДК 004.056

ПОИСК ОТСУТСТВИЙ СОБЫТИЙ В ЖУРНАЛЕ РЕГИСТРАЦИИ КАК СПОСОБ ВЫЯВЛЕНИЯ УЯЗВИМОСТЕЙ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

С. Л. Зефиров[1], А. Н. Аккуратнов[2]

[1,2] Пензенский государственный университет, г. Пенза

Аннотация. Определяется необходимость выявления фактов отсутствия событий в журналах. Рассматриваются недостатки текущих возможностей сигнатурного анализа для выявления отсутствующих событий. Предлагается оптимизированный метод поиска отсутствующих событий.

Ключевые слова: информационная безопасность, события информационной безопасности, требования документов информационной безопасности, уязвимость информационной безопасности, отсутствие событий

SEARCHING FOR ABSENCE OF EVENTS IN THE REGISTRATION LOG – AS A WAY TO IDENTIFY INFORMATION SECURITY VULNERABILITIES

S. L. Zefirov[1], A. N. Akkuratnov[2]

[1,2] Penza State University, Penza

Abstract. The need to identify the absence of events in the logs is determined. The shortcomings of the current capabilities of signature analysis for detecting missing events are considered. An optimized method for searching for missing events is proposed.

Keywords: information security, information security events, requirements for information security documents, information security vulnerability, no events

Международные и отечественные стандарты информационной безопасности определяют требования к системам мониторинга по регистрации и анализу событий информационной безопасности. Так, в разделе А.12.4 «Ведение журналов и мониторинг» основного стандарта по информационной безопасности *ISO 27001* [1] устанавливается следующее требование по регистрации событий:

«Должны вестись, сохраняться и регулярно анализироваться журналы, содержащие записи активности пользователей, возникновения исключений, сбоев и событий, связанных с информационной безопасностью».

© Зефиров С. Л., Аккуратнов А. Н., 2022

Требования к типам регистрируемых событий подробно устанавливаются в стандартах по обеспечению информационной безопасности финансовых организаций [2, 3]. Из довольно обширного списка типов регистрируемых событий можно выделить основные:

– административные действия (создание, удаление, изменение прав учетных записей);

– попытки логического доступа (в том числе и неуспешные);

– доступ ко всем журналам регистрации событий;

– факты выявления смены и (или) компрометации аутентификационных данных и т.д.

Цель регистрации событий четко определяется в стандарте ГОСТ Р ИСО/МЭК 27001–2006 [4] в разделе «А.10.10 Мониторинг»:

«Должны быть обеспечены ведение и хранение в течение определенного периода времени журналов аудита, регистрирующих действия пользователей, нештатные ситуации и события информационной безопасности, в целях помощи в будущих расследованиях и проведении мониторинга контроля доступа».

То есть, цель регистрации событий определяется как поиск инцидентов, несанкционированного доступа или аномальной активности. Само понятие «активность» определяет, что в журнале событий должны присутствовать события, характеризующие аномалию или несанкционированный доступ.

Эти же стандарты определяют требования к составу мер обеспечения информационной безопасности. Одной из таких мер является периодический анализ журналов событий. В п. 10.6 стандарта *PSI DSS* [2] требуется ежедневный анализ журналов всех серверов и системных компонентов, выполняющих основные технологические функции и функции безопасности. Аналогичные требования установлены в стандарте *NIST 800-92* [5]. Анализ журналов событий требует доступа к этим журналам. События доступа к журналам событий при проведении анализа должны ежедневно регистрироваться в журнале, исходя из требований по регистрации событий. Примерами соблюдения аналогичных требований, регистрируемых в журналах событий, могут служить плановая смена пароля пользователя, блокировка учетной записи пользователя после неуспешных попыток аутентификации, автоматическое прерывание сессии логического доступа по истечении установленного времени бездействия, выполнение периодических операций по проведению проверок на отсутствие вредоносного кода и т.д. Таким образом, анализируя журналы событий можно осуществлять контроль выполнения мер обеспечения информационной безопасности.

Если выполнение таких требований характеризуются наличием событий в журналах, то получается, что нарушение требований информационной безопасности характеризуется отсутствием событий. Отсутствие событий определенной категории в установленном интервале времени может указывать на то, что журналы ежедневно не анализируются, пароли не меняются, не проводятся проверки на отсутствие вредоносного кода и т.д. То есть создаются уязвимости информационной безопасности, использование которых может привести к инцидентам информационной безопасности. Поэтому для эффективного мониторинга важен не только поиск и анализ событий, характеризующих инциденты информационной безопасности, но и поиск отсутствующих событий, характеризующих выполнение мер обеспечения информационной безопасности.

В существующих системах мониторинга основным методом поиска инцидентов является сигнатурный анализ с использованием правил. Поиск отсутствующих событий (даже в реальном времени) обычно реализуется периодическими запросами к хранилищу событий и анализом количества возвращаемых событий в качестве результата. Если результат равен 0, то требуемое событие в журнале отсутствует. Поскольку интервалы между требуемыми событиями могут быть большими (например, плановая смена пароля должна осуществляться раз в месяц/квартал/год), то у такого подхода имеются недостатки. Требуется хранить большое количество данных в оперативном архиве, а также запросы с большим интервалом оказывают дополнительную нагрузку на производительность. Этот процесс можно оптимизировать. Поиск отсутствующих событий предлагается осуществлять использованием в процессе анализа переменной, значение которой содержит текущее время анализа, а в правилах анализа добавить возможность установки предельного времени появления события. Для анализа в реальном времени значением переменной текущего времени анализа может, например, являться текущее время сервера. Для проведения анализа на исторических данных начальное значение переменной устанавливается в значение начала указанного интервала анализа. Далее осуществляется приращение значения переменной с установленным шагом. На каждом шаге приращения осуществляется сравнение текущего времени анализа и предельного времени появления события, установленного в правиле. Если текущее время превышает предельное, то можно утверждать, что требуемое событие в журнале отсутствует.

Для установки предельного времени в правилах нужно определить виды отсутствия событий:

1. Относительное время (например, событие Б должно появиться после события А в течение 5 мин). В этом случае предельное время вычисляется как: время возникновения события А + 5 мин.

2. Абсолютное время (например, событие А должно появиться до 9:00 текущего дня). В этом случае предельное время устанавливается непосредственно при создании правила.

На рис. 1–3 показаны примеры правил с указанием временных характеристик для установки предельного времени появления событий по относительному и абсолютному времени. Для наглядности, правила показаны графически без привязки к конкретным системам мониторинга. Красным цветом выделены блоки, в которых устанавливаются временные параметры.

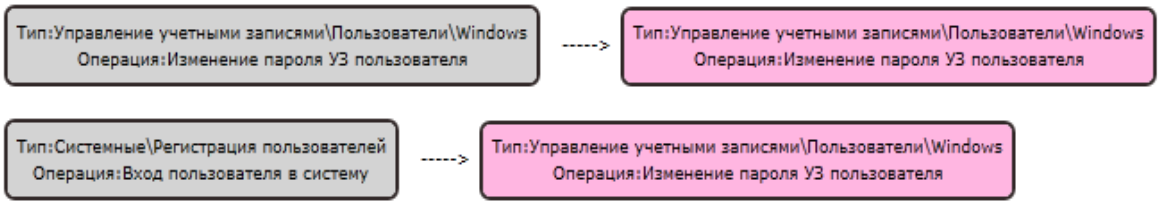

Рис. 1. Пример правила с установкой относительных временных параметров

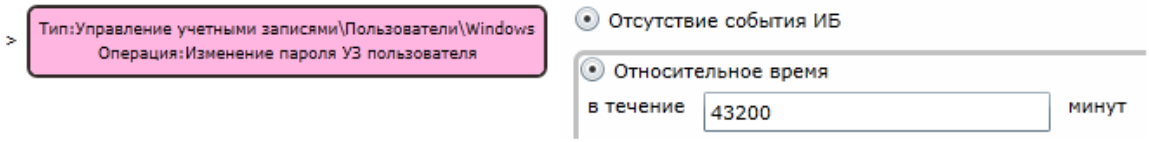

Рис. 2. Пример правила с установкой относительных временных параметров

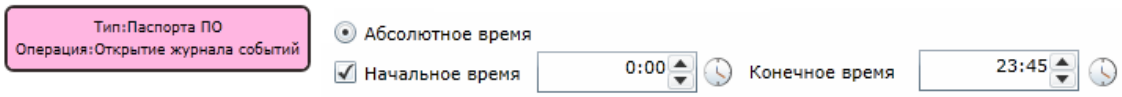

Рис. 3. Пример правила с установкой абсолютных временных параметров

На рис. 1 и 2 показан пример правила контроля выполнения требования по ежемесячной плановой смене пароля пользователя. Сначала выявляется событие смены пароля или активности пользователя. Далее к времени возникновения событий прибавляется 1 месяц (43200 мин) и устанавливается в качестве предельного времени появления события смены пароля. Если по истечению этого времени

событие смены пароля не будет найдено, значит можно сделать вывод, что пользователь пароль не менял и работает с просроченным паролем.

На рис. 3 показан пример правила контроля требования по ежедневному анализу журнала событий. Если в течение дня события доступа к журналу отсутствует, то анализ журнала не осуществлялся.

Предлагаемый метод поиска отсутствующих событий позволит организовать контроль соблюдения мер по обеспечению информационной безопасности и выявление уязвимостей информационной безопасности в случае их несоблюдения. Применение этого метода не потребует длительного хранения событий в оперативном архиве и запросов к нему за большие промежутки времени, что существенно снижает нагрузку на производительность любой системы мониторинга.

Список литературы

1. ISO/IEC 27001:2013. Информационные технологии. Методы защиты. Системы менеджмента информационной безопасности. Требования. 2013.

2. Payment Card Industry Data Security Standard (PCI DSS). Версия 3.2. 2016.

3. ГОСТ Р 57580.1–2017. Защита информации финансовых организаций. Базовый состав организационных и технических мер. 2017.

4. ГОСТ Р ИСО/МЭК 27001–2006. Методы и средства обеспечения безопасности. Системы менеджмента информационной безопасности. Требования. 2006.

5. NIST Special Publication 800-92. Guide to Computer Security Log Management. 2021.

6. NIST Special Publication 800-137. INFORMATION SECURITY. 2011.

Для цитирования: Зефиров С. Л., Аккуратнов А. Н. Поиск отсутствий событий в журнале регистрации как способ выявления уязвимостей информационной безопасности // Безопасность информационных технологий : сб. науч. ст. по материалам IV Всерос. науч.-техн. конф. : в 2 т. Пенза : Изд-во ПГУ, 2022. Т. 1. С. 97–101.

УДК 621.37.39

МЕТОДИКА ОБНАРУЖЕНИЯ КОМПЬЮТЕРНЫХ АТАК С ПОМОЩЬЮ ФРАКТАЛЬНОГО АНАЛИЗА И МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Р. А. Перов[1], С. Н. Ракицкий[2], С. В. Спирин[3], И. О. Евтихин[4]

[1,2,3,4] Военная академия связи имени Маршала Советского Союза С. М. Буденного Министерства обороны Российской Федерации, г. Санкт-Петербург

Аннотация. Рассмотрена методика выявления компьютерных атак для стационарного и нестационарного трафика. Проведен анализ алгоритмов машинного обучения для выявления аномалий, вызванных компьютерными атаками. Также вычислено минимальное количество пакетов для определения точности данной методики.

Ключевые слова: аномалии, компьютерные атаки, сеть передачи данных, метод машинного обучения, стационарный временной ряд, самоподобие

METHODS OF DETECTING COMPUTER ATTACKS USING FRACTAL ANALYSIS AND MACHINE LEARNING METHODS

R. A. Perov[1], S. N. Rakitsky[2], S. V. Spirin[3], I. O. Evtikhin[4]

[1,2,3,4] Military Academy of Communications named after Marshal of the Soviet Union S. M. Budyonny of the Ministry of Defense of the Russian Federation, St. Petersburg

Abstract. The article considers the method of detecting computer attacks for stationary and non-stationary traffic. The analysis of machine learning algorithms to identify anomalies caused by computer attacks is carried out. The minimum number of packets to determine the accuracy of this technique is also calculated.

Keywords: anomalies, computer attacks, data transmission network, machine learning method, stationary time series, self-similarity

Введение

С каждым годом увеличивается число компьютерных атак, направленных на крупные частные компании, государственные организации и критически важные информационные объекты.

Компьютерные атаки представляют собой сложное комплексное воздействие на сеть, в результате которого осуществляется их компрометация киберресурсов и нарушается управление процессами в сети передачи данных. Зачастую этому предшествует долгая

© Перов Р. А., Ракицкий С. Н., Спирин С. В., Евтихин И. О., 2022

и кропотливая работа: компьютерная техническая разведка, поиск характерных уязвимостей и захват информационных активов. Воздействие компьютерных атак возможно за счет использования технологий сбора информации, малоэффективных механизмов защиты, эксплуатации устаревших сетевых служб, протоколов и операционных систем.

Основная часть

С целью обнаружения аномалий в сетевом трафике в условиях компьютерных атак (КА) разработана методика (рис. 1), позволяющий обнаруживать компьютерные атаки, прогнозировать и обнаруживать компьютерные атаки с минимальным количеством ложных срабатываний благодаря применению методов фрактального анализа для нестационарного и методов машинного обучения для стационарного сетевого трафика [1].

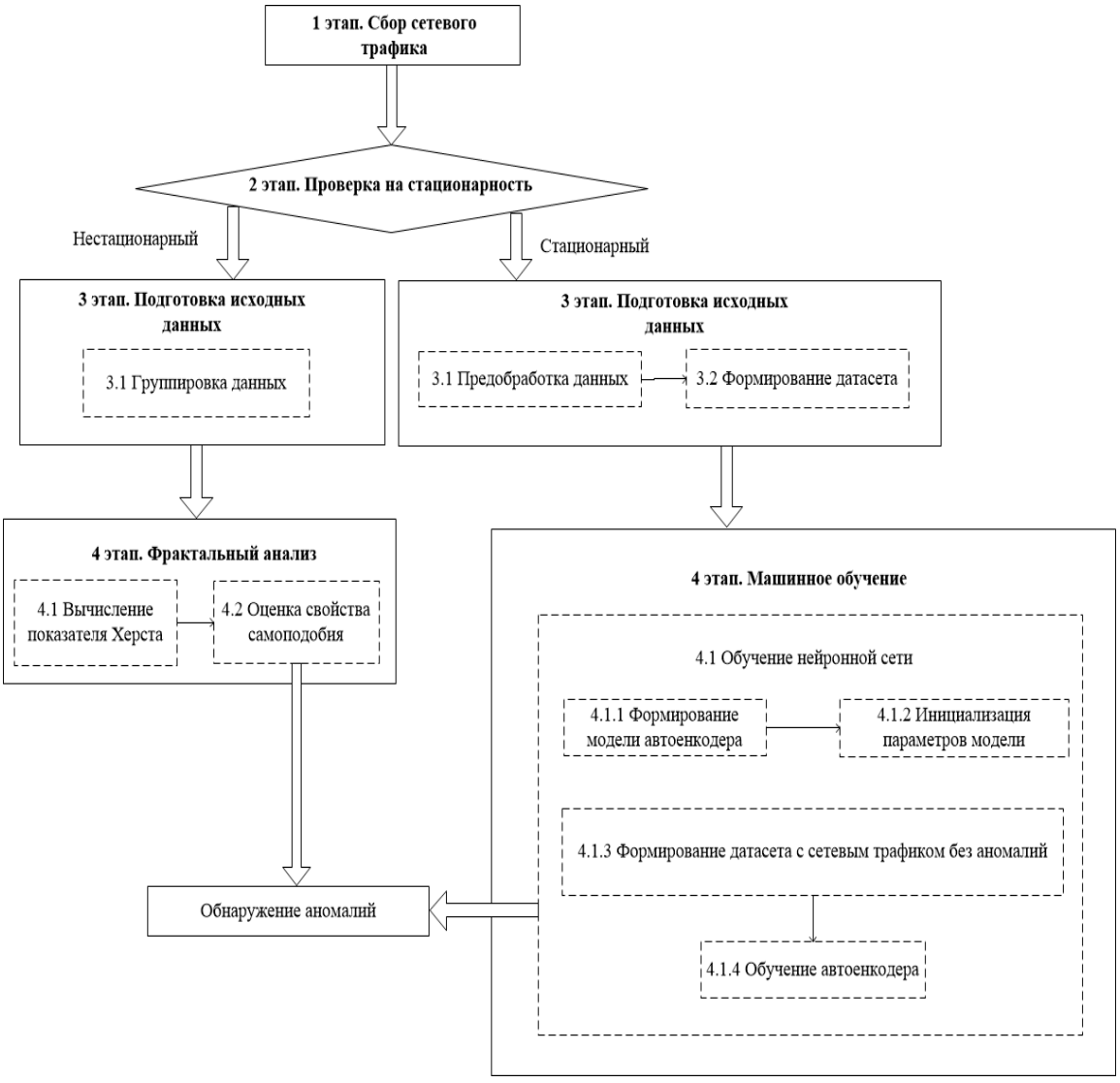

Рис. 1. Общая структура и взаимосвязь этапов метода

Методика включает следующие этапы:

– сбор сетевого трафика;

– проверка на стационарность;

– подготовка исходных данных;

– фрактальный анализ для нестационарного сетевого трафика;

– машинное обучение для стационарного сетевого трафика.

Обнаружение аномалий в нестационарном трафике с помощью фрактального анализа

Для выявления аномалий в нестационарном трафике, применяется фрактальный анализ, который главным образом базируется на вычислении и оценки показателя Херста.

Для оценки аномальности предложено использовать не полезное содержимое пакетов, а взять за основу предложение Унтерова Д. С. [2] о том, что информации из заголовков пакетов будет достаточно. В качестве такой информации будем использовать количественные значения передаваемых флагов (меток, указывающих на тип пакета).

Анализ [3] также показывает, что для выявления аномального поведения в трафике, достаточно анализировать его основные параметры и нет необходимости изучать содержимое каждого пакета. Примерами аномалий, обнаруженных на основе анализа телеметрии трафика, является внезапное увеличение интенсивности трафика от рабочей станции или изменение структуры в сравнении с обычными ежедневными показателями для данной сети устройства.

На рис. 2 представлена блок-схема метода обнаружения аномальной активности в нестационарном сетевом трафике сети передачи данных (СПД).

Сетевой поток делится на группы и рассчитывается показатель Херста для каждой из групп. Сетевые пакеты помечаются аномальными в том случае, когда нарушается свойство самоподобия в исследуемой группе.

Для проверки эффективности предложенного подхода, сперва он был протестирован на легитимном сетевом трафике (рис. 3).

Синей прямой линией обозначен порог, соответствующий границе белого шума (Hurst = 0,5). Точки на втором графике соответствуют номерам групп пакетов (всего 30 точек). Точки на третьем соответствуют scales (всего 12 точек). Количество scales влияет на точность и на длительность работы алгоритма. Чем больше количество scales, тем выше точность, и, наоборот, меньше длительность его работы.

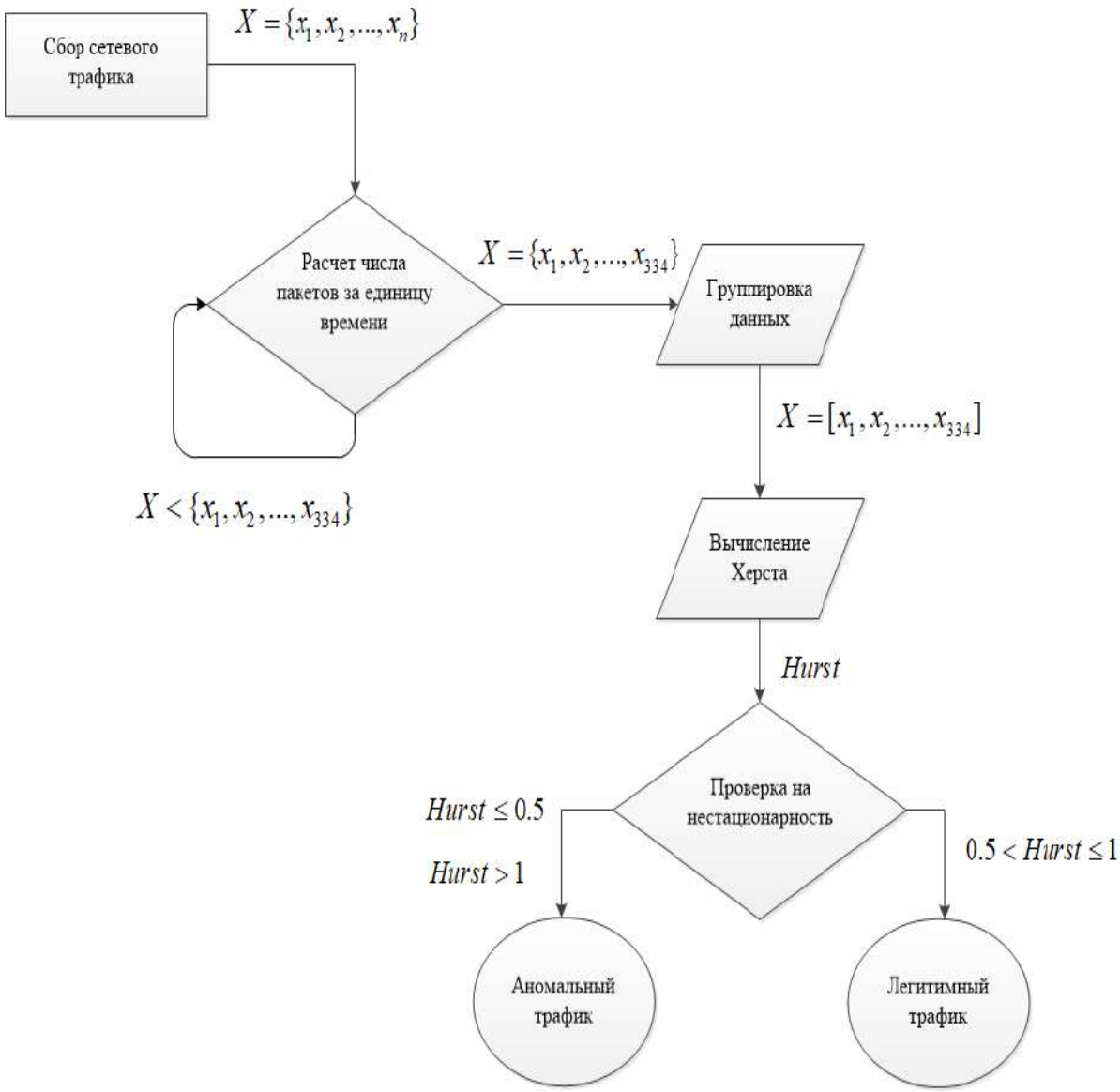

Рис. 2. Блок-схема обнаружения аномалий
в нестационарном трафике

Как видно из рис. 4, мера фрактальности для всех групп пакетов полностью лежит выше отметки 0,5. Это указывает на наличие самоподобных свойств у каждой группы. Кроме того, на третьем графике (логарифмической регрессии) отражен параметр Херста для всего DataFrame, который подтверждает наличие фрактальных свойств и повторяющихся процессов.

Далее проводилось тестирование аномального сетевого трафика, полученного во время проведения DoS атаки и компьютерно-технической разведки. При этом преследовалась цель подбора максимального числа групп разбиения, при котором оцениваемый параметр H будет вычисляться с высокой точностью.

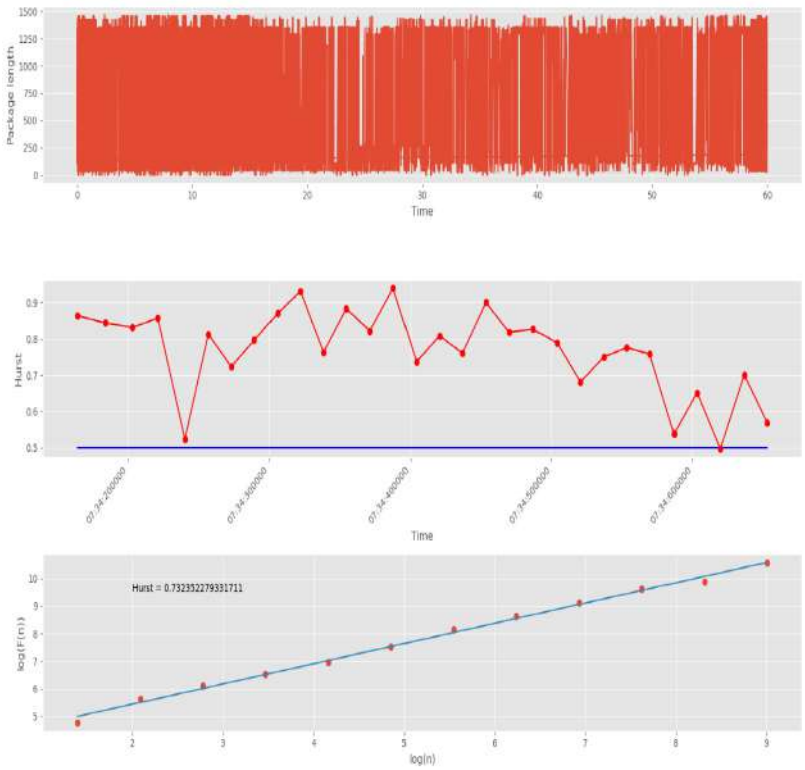

Рис. 3. Вычисление H методом фрактального анализа легитимного UDP трафика. Разбиение 10000 точек на 20 групп

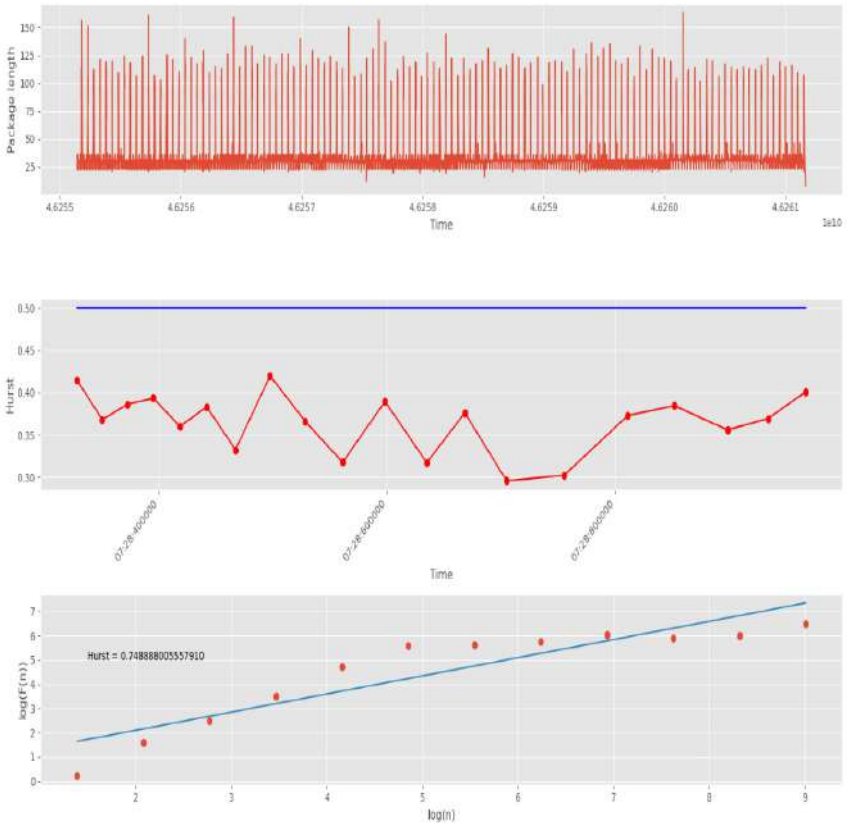

Рис. 4. Вычисление H методом фрактального анализа аномального UDP трафика. Разбиение 10000 точек на 20 групп

Из рис. 4 видно, что свойство самоподобия нарушается, так как показатель Херста, на каждом из интервалов, меньше порогового значения 0,5. Это свидетельствует о нарушении фрактальной структуры трафика и наличии в нем аномалий. Следовательно, предлагаемый подход способен обнаруживать аномалии в интервалах (группах), состоящих из 500 сетевых пакетов.

Увеличив количество групп до 30, сокращается временной интервал без потери точности обнаружения аномальной активности в сети (рис. 5).

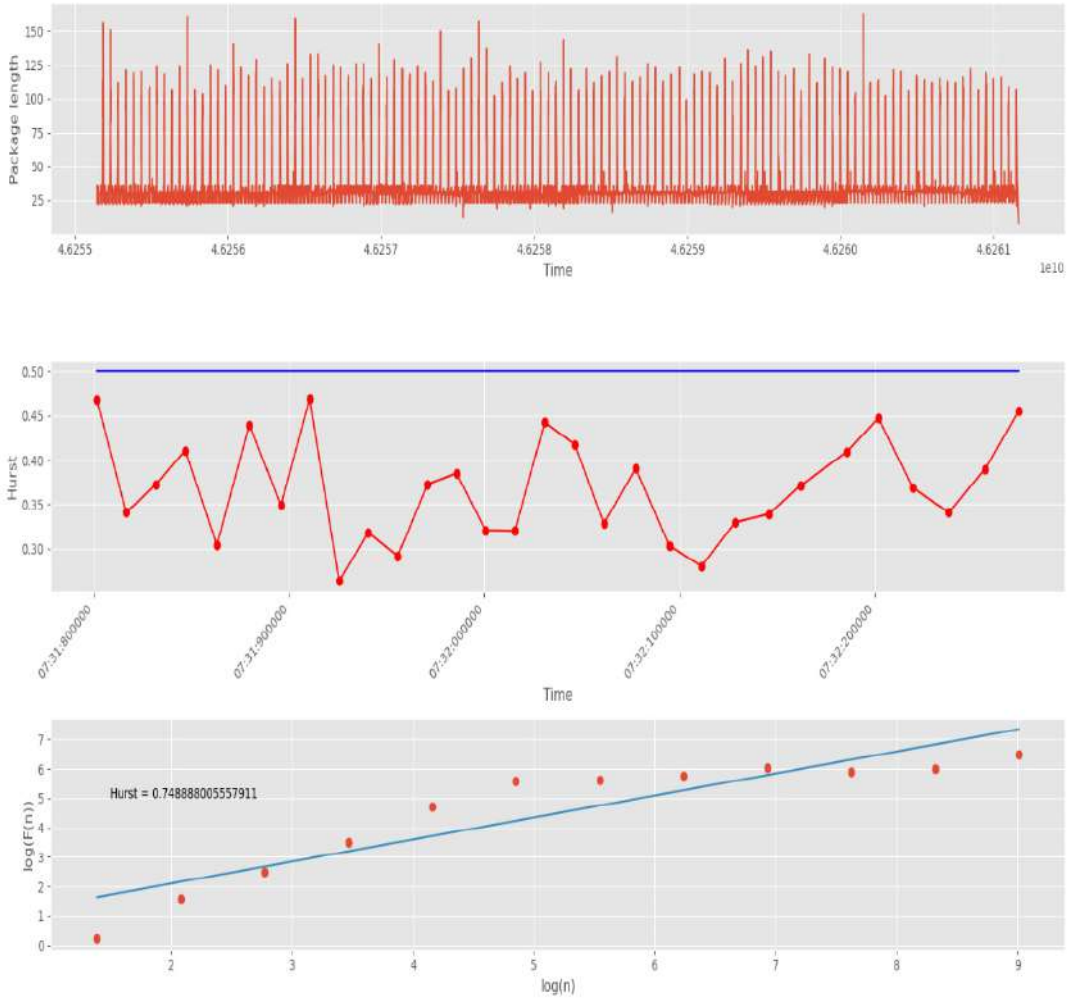

Рис. 5. Вычисление H методом фрактального анализа аномального UDP трафика. Разбиение 10000 точек на 30 групп

При анализе 40 групп, каждая из которых состоит из 250 сетевых пакетов, наблюдается проявление свойств самоподобия на некоторых участках (рис. 6).

Такое поведение указывает на ухудшение точности из-за малой выборки сетевых пакетов. Следовательно, такое разбиение является неприемлемым. В качестве оптимального разбиения следует считать

предыдущее разбиение, состоящее из 250 сетевых пакетов на интервал. Такое количество пакетов обрабатывается за 0,00125 с, что является существенным достоинством данного подхода.

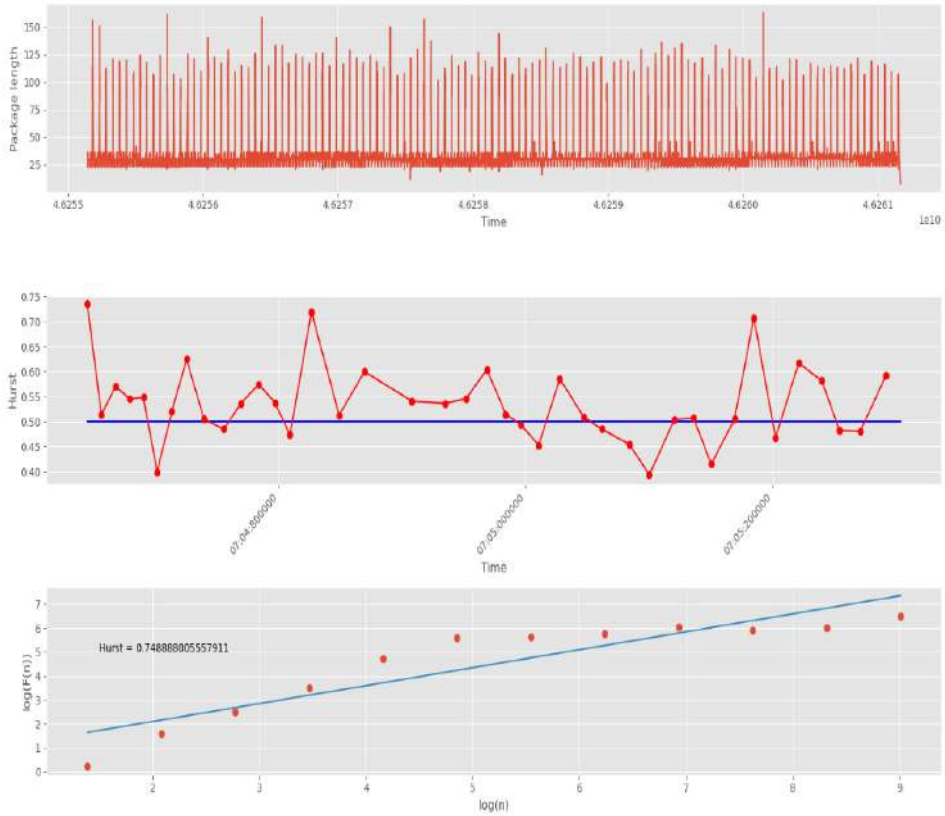

Рис. 6. Вычисление H методом фрактального анализа аномального UDP трафика. Разбиение 10000 точек на 40 групп

Обнаружение аномалий в стационарном сетевом трафике с помощью методов машинного обучения

Существует множество способов, которые позволяют определить аномалии [4]. На рис. 7–9 продемонстрирована работа наиболее популярных алгоритмов машинного обучения, протестированных на временных рядах, сгенерированных с помощью модели авторегрессионного интегрированного скользящего среднего.

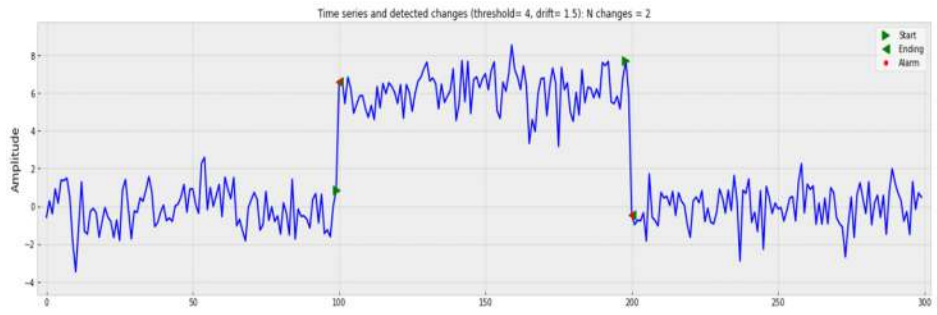

Рис. 7. Кумулятивные суммы. Начало

Рис. 7. Окончание

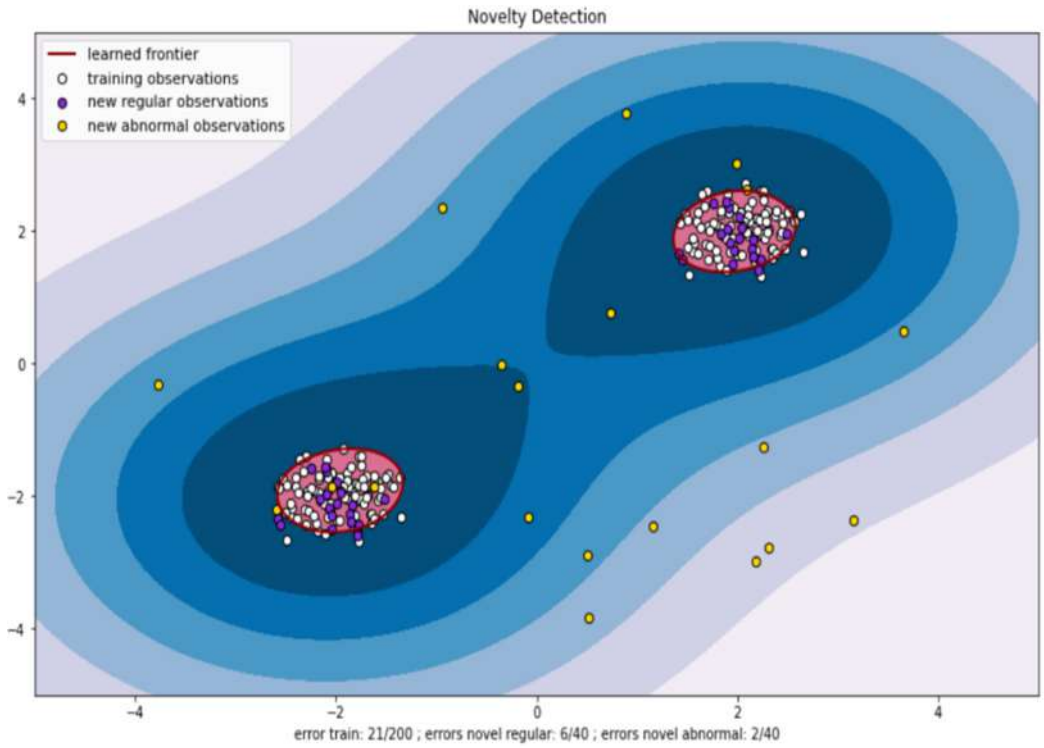

Рис. 8. Метод опорных векторов

Как видно из рисунков, алгоритмы отлично справляются с обнаружением аномальных выбросов. В таком случае, аномалия проявляется в виде нестационарности некоторых наблюдаемых временных

рядов. Это не только мгновенные скачки амплитуды измерений, но и медленные тренды, практически невидимые за время наблюдений. Однако, при тестировании вышеуказанных алгоритмов на реальном сетевом трафике, оказалось, что не всегда выбросы являются аномальными.

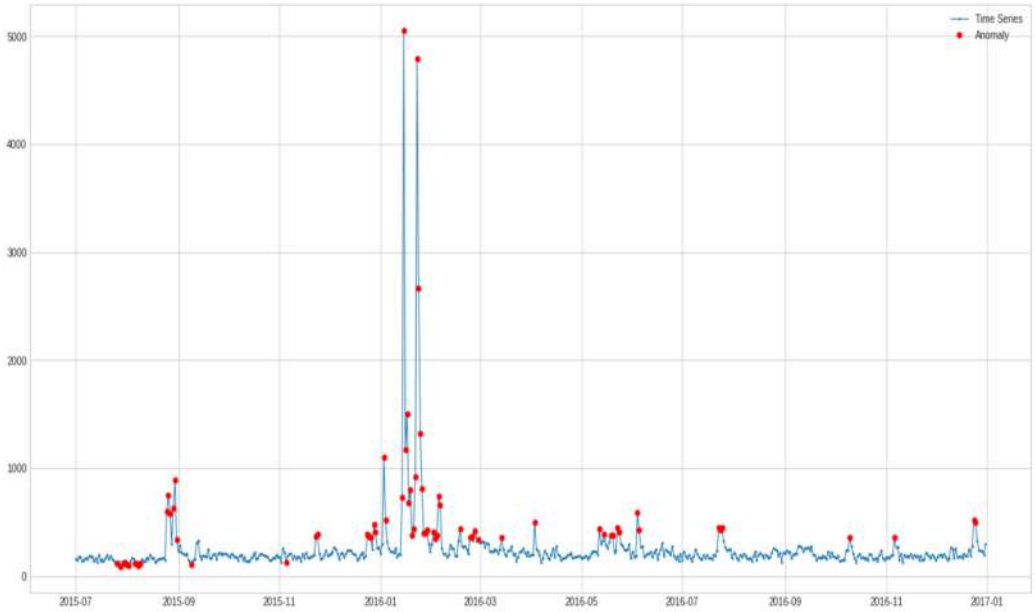

Рис. 9. Изолированный лес

На рис. 10–13 изображены протоколы сетевого трафика. Аномальные пакеты помечены красными точками, а легитимные пакеты – зелеными. Как видно из рисунков многие всплески являются легитимными, а отсутствующие всплески – аномальные.

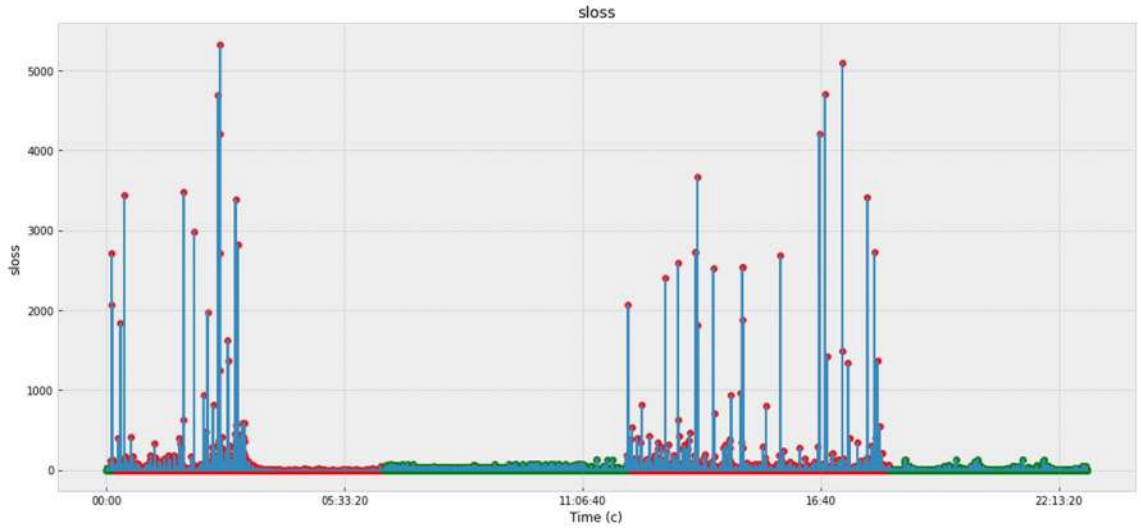

Рис. 10. Отправленные пакеты повторно переданы или отброшены

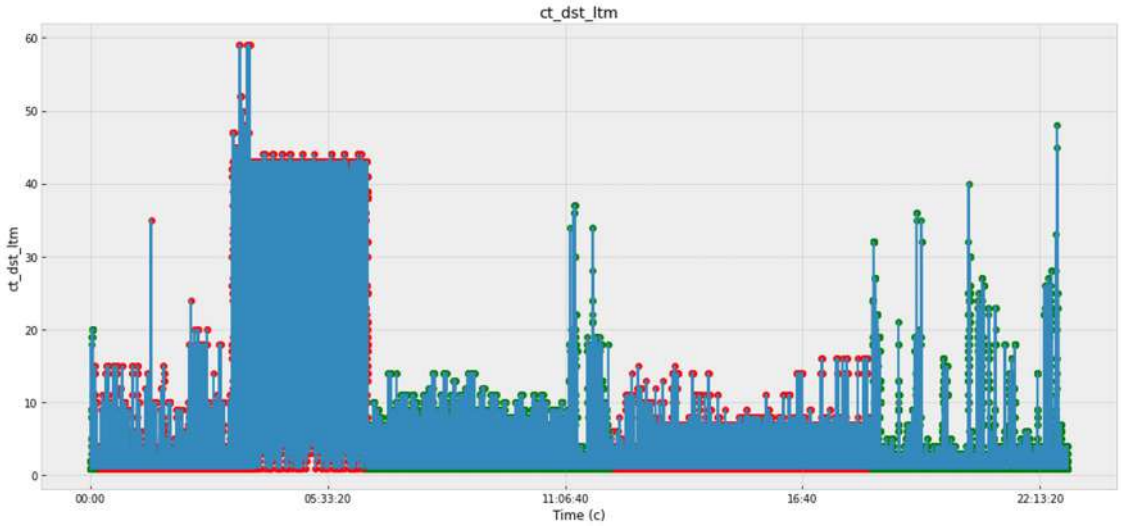

Рис. 11. Количество подключений к серверу

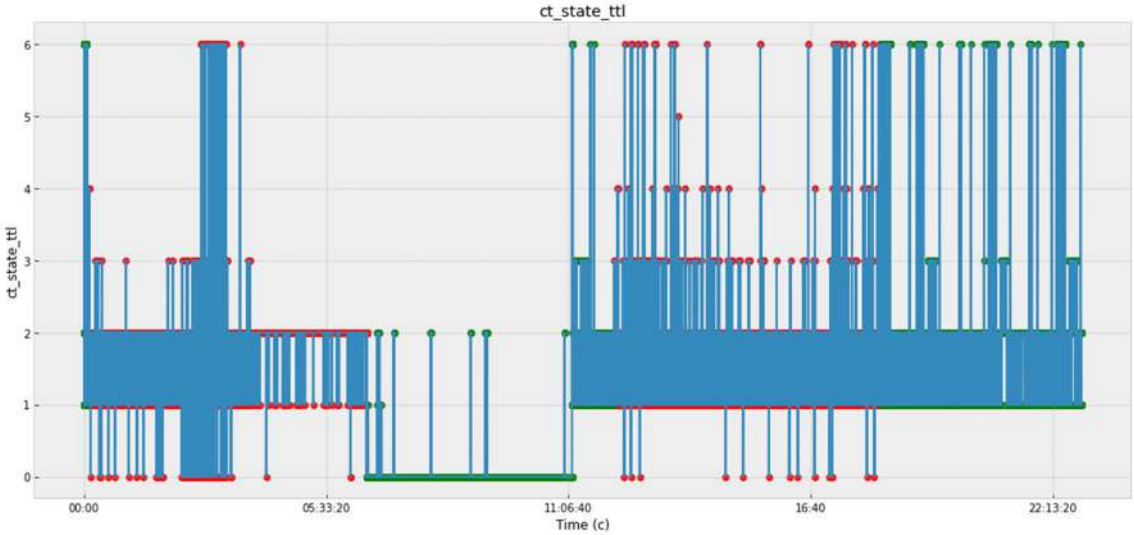

Рис. 12. Состояние параметров tcp заголовка за время жизни ip пакета

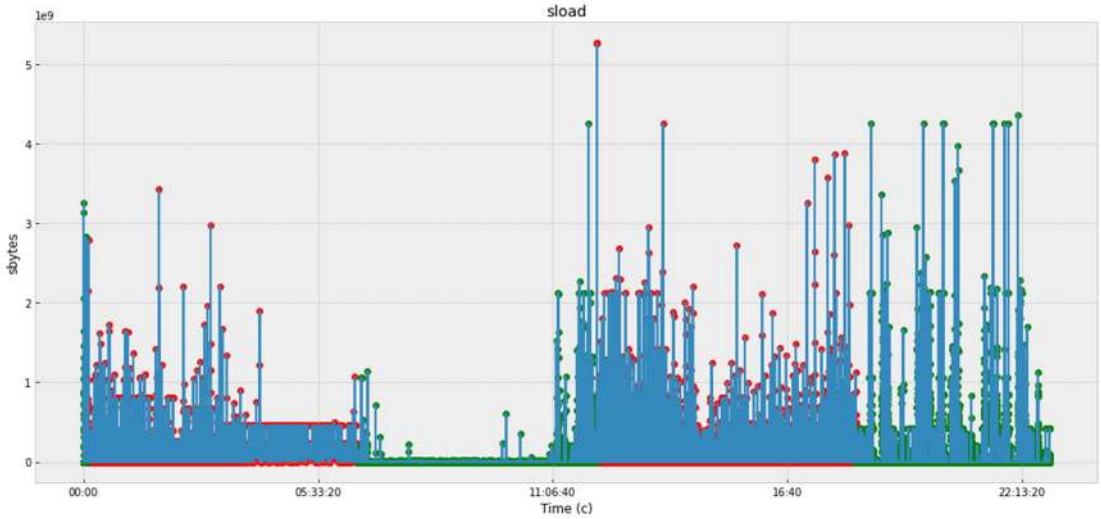

Рис. 13. Скорость передачи пакетов bit/c

Поэтому остро стоит вопрос по своевременному обнаружению всплесков трафика в СПД, выделению из них аномальных, а также классификации выявленных аномалий с целью прогнозирования факта воздействия КА и выработке эффективных мероприятий противодействия.

Полученные результаты подтверждают тот факт, что проблемой современных методов машинного обучения является плохая способность распознавать неизвестные аномалии (атаки нулевого дня). Также, при тестировании вышеуказанных алгоритмов на реальном сетевом трафике оказалось, что не всегда выбросы являются аномальными; происходят ложные срабатывания. Поэтому острым является вопрос классификации компьютерных атак от ложных срабатываний.

Заключение

В статье разработана методика, позволяющая выявлять аномалии в нестационарном сетевом трафике с помощью фрактального анализа. Проведен эксперимент по нахождению рационального числа пакетов, необходимых для своевременного выявления аномалий с помощью фрактального анализа, с целью достижения своевременности и полноты при обнаружении аномалий в нестационарном трафике. Вычислено оптимальное число пакетов необходимых для выявления аномалий в СПД за интервал времени.

Проведен анализ алгоритмов машинного обучения и классификаторов, предназначенных для выявления КА. Выявлены их недостатки, которые главным образом заключаются в большом количестве ложных срабатываний, а также отсутствии возможности работать в режиме реального времени и обнаружения КА «нулевого дня».

Главным достоинством методики является возможность работы в режиме реального времени, а также возможность работы с любыми видами трафика. Выявление факта воздействия КА производится за несколько микросекунд в зависимости от производительности вычислительной техники.

Список литературы

1. Kotenko I., Saenko I., Lauta O., Kribel A. Ensuring the survivability of embedded computer networks based on early detection of cyber-attacks by integrating fractal analysis and statistical methods // Microprocessors and Microsystemsthis link is disabled. 2022. Vol. 90. P. 104459.

2. Саенко И. Б., Лаута О. С., Карпов М. А., Крибель А. М. Модель угроз ресурсам ИТКС как ключевому активу критически важного объекта инфраструктуры // Электросвязь. 2021. № 1. С. 36–44.

3. Карпов М. А., Коцыняк М. А., Нечепуренко А. П. Модель функционирования информационно-телекоммуникационной сети специального назначения в условиях информационного воздействия // Актуальные проблемы защиты и безопасности : тр. XXIV Всерос. науч.-практ. конф. РАРАН : в 7 т. (г. Санкт-Петербург, 31 марта – 3 апреля 2021 г.). М. : Российская академия ракетных и артиллерийских наук, 2021. С. 458–462.

4. Лепешкин О. М., Карпов М. А., Остроумов О. А., Синюк А. Д. Методологический подход управления обеспечением функциональной безопасности и функциональной устойчивости системы связи критически важных объектов и объектов критической информационной инфраструктуры // FISP-2021: Фундаментальные проблемы информационной безопасности в условиях цифровой трансформации : сб. докл. III Всерос. науч. конф. (с приглашением зарубежных ученых). (г. Ставрополь, 30 ноября 2021 г.). Ставрополь : Северо-Кавказский федеральный ун-т, 2021. С. 105–110.

Для цитирования: Перов Р. А., Ракицкий С. Н., Спирин С. В., Евтихин И. О. Методика обнаружения компьютерных атак с помощью факториального анализа и методов машинного обучения // Безопасность информационных технологий : сб. науч. ст. по материалам IV Всерос. науч.-техн. конф. : в 2 т. Пенза : Изд-во ПГУ, 2022. Т. 1. С. 102–113.

УДК 621.398

МЕТОДИКА ПОВЫШЕНИЯ ЗАЩИЩЕННОСТИ СЕТЕЙ ПЕРЕДАЧИ ДАННЫХ НА ОСНОВЕ МЕТОДА КОНТРОЛЯ УЯЗВИМОСТЕЙ

В. Б. Сычужников[1], С. В. Спирин[2], К. Р. Бакмаева[3], И. О. Евтихин[4]

[1,2,3,4] Военная академия связи имени Маршала Советского Союза С. М. Буденного Министерства обороны Российской Федерации, г. Санкт-Петербург

Аннотация. Рассмотрен подход к контролю неизменности параметров безопасности средств вычислительной техники, являющийся составной частью процесса восстановления безопасного состояния сети передачи данных в совокупности трех событий, происходящих в случайные моменты времени: непосредственно контроль неизменности параметров безопасности, санкционированное изменение, реализация угроз параметров безопасности. Разработана частная методика расчета рациональных параметров контроля состояния их неизменности, выходные данные которой могут использоваться для настройки специального программного обеспечения по выполнению контроля.

Ключевые слова: сеть передачи данных, безопасность связи и информации, программное обеспечение, контролируемые параметры, файлы конфигурации, уязвимости сетей

METHODS FOR IMPROVING THE SECURITY OF DATA TRANSMISSION NETWORK BASED ON THE VULNERABILITY CONTROL METHOD

V. B. Sychuzhnikov[1], S. V. Spirin[2], K. R. Bakmayeva[3], I. O. Evtikhin[4]

[1,2,3,4] Military Academy of Communications named after Marshal of the Soviet Union S. M. Budyonny of the Ministry of Defense of the Russian Federation, St. Petersburg

Abstract. An approach to the control of the immutability of the security parameters of computer equipment is considered, which is an integral part of the process of restoring the safe state of the data transmission network in the aggregate of three events occurring at random moments of time: direct control of the immutability of the security parameters, authorized change, implementation of threats to the security parameters. A particular method for calculating rational parameters of monitoring the state of their immutability has been developed, the output data of which can be used to configure special software for monitoring.

Keywords: data transmission network, communication and information security, software, controlled parameters, configuration files, network vulnerabilities

© Сычужников В. Б., Спирин С. В., Бакмаева К. Р., Евтихин И. О., 2022

Введение

Сети передачи данных (СПД), составляющие часть сетей связи специального назначения (СС СН), являющиеся по сути информационно-телекоммуникационными сетями (ИТКС) [1], выполняют задачи по обеспечению системы управления войсками достоверной и своевременной информацией для выполнения задач управления. Основной особенностью СС СН, которая отличает их от сетей связи общего пользования (СС ОП), является то, что СС СН ориентированы на функционирование как в мирное, так и в военное время, в условиях воздействия противника, а также различного рода дестабилизирующих факторов. В связи с этим для СС СН особенное значение приобретает свойство их безопасности.

В настоящее время при недостатке собственных ресурсов СС СН, как правило, необходимые канальные ресурсы из СС ОП арендуются у региональных и национальных операторов связи. Из самого факта сопряжения СС СН и СС ОП следует два важных вывода: 1) технологии связи СС СН должны быть «обратно совместимыми» с технологиями, используемыми в гражданских СС ОП, для обеспечения использования ресурса СС ОП в интересах СС СН; 2) сквозное сопряжение СС СН с СС ОП, а также последней с гражданскими СС ОП других государств, делает СС СН потенциально уязвимыми для информационно-технических воздействий со стороны других государств [2].

Основная часть

В процессе эксплуатации СПД, средствами вычислительной техники (СВТ) формируется и периодически обновляется информация об их состоянии, составляющая контролируемые параметры (КП) СВТ, являющаяся критичной для безопасности связи и информации (БСИ). К такой информации могут относиться конфигурации систем управления, фильтрации файрволлов, параметры адресации маршрутизаторов и прочие. Умышленные действия оператора СВТ по изменению КП могут быть квалифицированы как неразрешенные (небезопасные с точки зрения информации и связи). Данные действия сопутствуют возникновению уязвимостей СПД.

Источники [3, 4] содержат обширный круг угроз и уязвимостей, в результате реализации которых обрабатываемые (хранимые, передаваемые) данные могут быть похищены, искажены или уничтожены. В совокупности указанные факты определяют актуальность обеспечения неизменности параметров безопасности СВТ. Под неизменностью параметров безопасности понимается такое их состояние, при

котором любое изменение отсутствует либо выполняется преднамеренно уполномоченными на это субъектами.

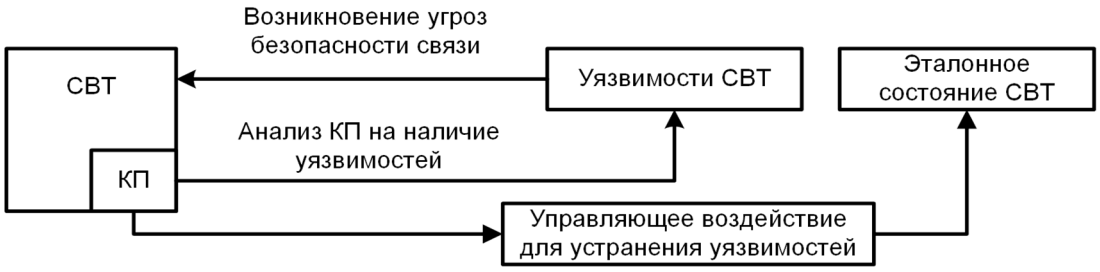

Рис. 1 Концептуальное представление методики повышения защищенности сетей передачи данных на основе метода поиска уязвимостей

Под уязвимостью (брешью, изъяном в защите) информационной системы понимают свойство информационной системы, представляющее возможность реализации угроз безопасности, обрабатываемой в ней информации [5].

В качестве примера удобно рассмотреть известную уязвимость BDU:2020-00877. Это уязвимость веб-интерфейса управления систем обеспечения безопасности электронной почты. Эксплуатация уязвимости может позволить нарушителю, действующему удалённо, выполнить произвольный код в контексте текущего пользователя или раскрыть защищаемую информацию с помощью специально созданной ссылки. Данная уязвимость относится к производителю Cisco Systems Inc. и обнаружена в программном обеспечении (ПО) Cisco Email Security Appliances. Уязвимость имеет средний уровень опасности и встречается в ПО Cisco Email Security Appliances версии 13.0 и ранее; способом устранения уязвимости является обновление версии ПО.

Другим примером уязвимости является возможность передачи учетных данных с использованием незащищенных сетевых протоколов. К ним относится, например, протокол HTTP. Злоумышленник, перехватывая и анализируя передаваемый трафик, может извлечь из него передаваемые в сеть логины и пароли. Чаще всего, возможность применения незащищенных сетевых протоколов при передаче информации, является настраиваемым параметром прикладного ПО (браузера).

Третьим примером уязвимости является возможность использования незащищенных протоколов удаленного управления. В настоящее время, существует несколько незащищенных протоколов удаленного управления, к числу которых можно отнести протокол Telnet. При использовании данного протокола, аутентификация

выполняется на базе открытого текста (пароли пересылаются в незашифрованном виде). Угроза применения протоколов удаленного управления аналогична угрозе передачи учетных данных с использованием незащищенных сетевых протоколов и может грозить извлечением учетных записей пользователей и получением удаленного доступа к хостам. Чаще всего, возможность использования незащищенных сетевых протоколов, является настраиваемым параметром программно-аппаратных средств защиты информации (криптомаршрутизаторов) [6].

Для повышения защищенности СПД при использовании СВТ применяется метод поиска уязвимостей. Поиск уязвимостей представляет собой бинарную проверку КП на соответствие эталону. Он должен выполняться таким образом, чтобы до момента времени потенциальной реализации угрозы БСИ, с требуемой вероятностью была выявлена возникшая уязвимость, после чего оказано управляющее воздействие, позволяющее в минимальные сроки осуществить ее устранение.

Алгоритм, направленный на поиск уязвимостей СВТ, представляет следующую последовательность действий.

1. Определение совокупности подконтрольных СВТ.

2. Выбор множества параметров, необходимых для контроля (критичных с точки зрения БСИ).

3. Определение допустимых границ изменения КП.

4. Определение промежуточных временных интервалов контроля параметров безопасности.

Пусть имеется СПД с объемом контролируемых параметров $V_{\text{КП}}$, составляющих ее защищенность, и корректность которых критична для безопасного состояния. Эти параметры периодически изменяются уполномоченным должностным лицом (оператором) СПД, время изменения $T_{\text{изм}}$ является случайной величиной и фактически определяется предназначением СВТ и режимом его работы. Будем считать, что значение периодичности санкционированных изменений $t_{\text{изм}}$ подчинено нормальному закону распределения с параметрами средней периодичности изменений $T_{\text{изм.КП}}$ и средним квадратическим отклонением $\sigma_{\text{изм.КП}}$. Плотность распределения вероятностей для нормального закона описывается выражением:

$$f(t_{\text{изм}} = \frac{1}{\sigma_{\text{изм.КП}} \cdot \sqrt{2\pi}} e^{\frac{(t_{\text{изм}} - T_{\text{изм.КП}})^2}{2 \cdot \sigma_{\text{изм.КП}}^2}} . \quad (1)$$

На состояние КП оказывает воздействие множество внешних и внутренних факторов, представляющих угрозы БСИ $У = \left\{ y_i \mid i = \overline{1, N_{\text{угр}}} \right\}$, где $N_{\text{угр}}$ – количество факторов возможных угроз. В результате реализации любой из угроз происходит изменение КП, защищенность снижается, в результате чего состояние БС нарушается. Так как проявление какого-либо фактора к моменту времени *t* является случайным событием, обозначим случайной величиной $T_{\text{угр } i}$ – время до реализации угрозы целостности КП *i*-го фактора. Если потоки угроз *i*-ого фактора простейшие, то интенсивность угроз *i*-го фактора $\lambda_{\text{угр } i} = 1 / T_{\text{угр } i}$. Суммарный поток угроз также является простейшим [7], интенсивность суммарного потока угроз $\lambda_{\text{угр } i}$ и среднее время до реализации любой из угроз $T_{\text{угр}}$:

$$\lambda_{\text{угр}} = \sum\nolimits_{i=1}^{N_{\text{угр}}} \lambda_{\text{угр } i} = \sum\nolimits_{i=1}^{N_{\text{угр}}} T_{\text{угр } i}^{-1}, \tag{2}$$

$$T_{\text{угр}} = \lambda_{\text{угр}}^{-1} = \frac{1}{\sum\nolimits_{i=1}^{N_{\text{угр}}} T_{\text{угр } i}^{-1}}. \tag{3}$$

Допущение о простейшем потоке угроз при рассмотрении моделей безопасности является широко распространенным и применяется при решении большого количества задач [8–10]. Оно подразумевает, что поток изменений КП является стационарным, ординарным и без последействия.

На современном этапе развития систем контроля БС и ее нормативно-правового регулирования целесообразно разрабатывать методический аппарат на основе эмпирических данных об интенсивности суммарного потока угроз $\lambda_{\text{угр}}$ или о среднем времени до реализации любой из угроз $T_{\text{угр}}$ [11].

Исходные данные разрабатываемой методики обеспечения своевременности выявления уязвимостей СПД представлены в табл. 1.

Выходными данными методики будут параметры согласно табл. 2.

Определение вида контроля *w* предполагается из следующего множества $W = \left\{ w_i \mid i = 0..1 \right\}$ доступных вариантов: w_0 – полный контроль состояния СВТ, w_1 – рациональный контроль состояния КП СВТ. Под рациональным контролем понимается не контроль параметров безопасности СВТ (требующий привлечения значительных ресурсов контроля), а контроль неизменности файлов, содержащих параметры конфигураций контролируемых СВТ. Хеш-суммы файлов конфигураций (ФК), рассчитываются и фиксируются после полного контроля безопасности СВТ.

Таблица 1

Исходные данные методики повышения защищенности СПД на основе метода поиска уязвимостей

№ п/п	Наименование исходных данных	Обозначение
1	Среднее время до возникновения угрозы	$T_{\text{угр}}$
2	Требуемая вероятность обеспечения своевременности контроля	$P_{\text{треб}}$
3	Объем базы данных контролируемых параметров	$V_{\text{КП}}$
Сведения о частоте изменения информации		
4.1	Наличие достоверной информации о факте изменения КП в момент времени *t* по отношению к моменту последнего измерения	$I_{\text{изм}}(t)$
или		
4.2	Параметры закона распределения времени до очередного изменения информации	$T_{\text{изм}}, \sigma_{\text{изм.КП}}$

Таблица 2

Выходные данные методики повышения защищенности СПД на основе метода поиска уязвимостей

№ п/п	Наименование исходных данных	Обозначение
1	Время до очередного контроля	$t_{\text{контр}}$
2	Вид контроля	$w \in W = \{w_i \mid i = 0..1\}$

Основными допущениями и ограничениями методики в основном являются ограничения и допущения, принятые в модели эксплуатации СВТ. При этом если сведения о факте изменения КП $I_{\text{изм}}(t)$ в момент времени *t* отсутствуют, то предполагается, что изменения КП осуществляются по нормальному закону (в качестве исходных данных задан п. 4.2 вместо п. 4.1 табл. 1).

Следует отметить, что в настоящее время существует немало программных продуктов, которые способны выполнять контроль целостности ФК параметров безопасности СВТ требуемым, в установленные промежутки времени. К ним, например, относятся: Zabbix, MaxPatrol SIEM, Rancid, Nagios, Cacti, Munin, Centreon, Sensu, PolyMon, PowerAdmin и др. В случае использования какого-либо вспомогательного ПО для контроля, осуществляется его настройка согласно полученных из частной методики параметрам.

Исходя из целей, задач и сущности разрабатываемой методики для обеспечения своевременного выявления уязвимостей она должна применяться на протяжении всего этапа эксплуатации к каждому отдельному СВТ.

Рациональный контроль безопасности контролируемых параметров.

Рациональная периодичность контроля, т.е. моменты времени, когда необходимо выполнять очередной опрос целостности файлов КП СВТ и вид контроля, зависит от двух событий: вероятности наступления какой-либо угрозы КП, приводящей к нарушению БСИ и осуществления изменения информации уполномоченным оператором (возникновения уязвимости). Оба этих событий являются случайными, а время их наступления – случайной величиной.

Обобщенная блок-схема последовательности действий по расчету рациональных параметров контроля, которые составляют частную методику, представлена на рис. 5.

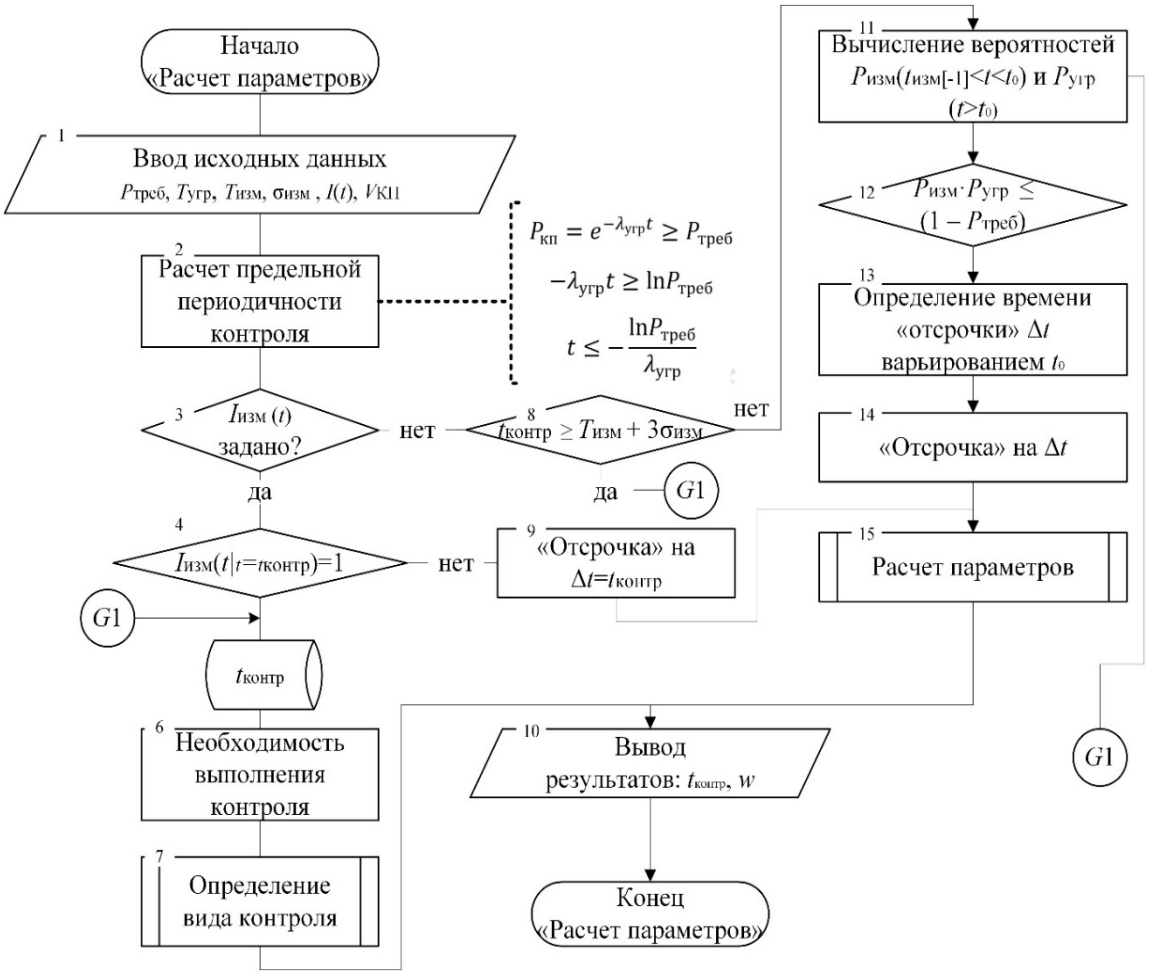

Рис. 2. Блок-схема частной методики расчета рациональных параметров контроля

В блоке 1 задаются исходные данные.

Полагая, что время до возникновения угроз изменения КП $T_{КП}$, являясь случайной величиной, описывается экспоненциальным

законом распределения с параметром среднего времени $T_{\text{угр}}$, вероятность «не изменения КП» $P_{\text{КП}}(t)$ определяется:

$$P_{\text{КП}}(t) = P(T_{\text{КП}} > t) = e^{-\lambda_{\text{угр}} t} = e^{\frac{-t}{T_{\text{угр}}}} \quad (4)$$

Исходя из требуемого условия обеспечения целостности ФК с требуемой вероятностью

$$P_{\text{КП}}(t) = e^{\frac{-t}{T_{\text{угр}}}} \geq P_{\text{треб}} \quad (5)$$

находим время

$$-\frac{t}{T_{\text{угр}}} \geq \ln P_{\text{треб}}; \quad (6)$$

$$t = t_{\text{контр}} \leq -T_{\text{угр}} \cdot \ln P_{\text{треб}}. \quad (7)$$

Значение времени $t_{\text{контр}}$ (блок 2), вычисленное согласно выражению (7), является верхним пределом, ограничивающим максимальную периодичность контроля для исключения возможности эксплуатации уязвимости противником с требуемой достоверностью.

Исключение ситуаций, когда к рассчитанному по (7) моменту времени информация о возникновении уязвимости отсутствует, зависит от характера сведений о вероятности изменения КП (возникновении уязвимости) – детерминированного или вероятностного, т.е. от способа задания исходных данных (блок 3). В первом случае (п. 4.1 исходных данных таблицы 1) подразумевается, что располагаем достоверными сведениями $I_{\text{изм}}(t)$ о наличии или отсутствии данных об изменении КП за период $\left(t_{\text{контр}[-1]}; t\right)$, т.е. с момента предыдущего контроля $t_{\text{контр}[-1]}$. Формат таких сведений имеет следующий вид:

$$I_{\text{изм}}(t) = \left\{ \begin{array}{l} 1, \text{ если } t_{\text{контр}[-1]} \leq t_{\text{изм}} < t \\ 0, \text{при}(t_{\text{изм}} < t_{\text{контр}[-1]}) OR (t_{\text{изм}} > t) \end{array} . \right\} \quad (8)$$

Если изменения в указанном периоде были, т.е. $I_{\text{изм}}(t) = 1$ (блок 4), то полученное значение $t_{\text{контр}}$ фиксируется (блок 5). Очередной контроль в момент времени $t_{\text{контр}}$ необходимо выполнить (блок 6). В противном случае вернуться к контролю через время $\Delta t = t_{\text{контр}}$

(блок 9). Такая «отсрочка» в выполнении контроля не приведет к превышению вероятности эксплуатации уязвимости за время Δt над требуемой вероятностью обеспечения БС и обоснована принятым допущением об экспоненциальном законе потока угроз КП.

Когда сведения о фактах и времени возникновения угроз КП отсутствуют, приходится учитывать вероятностный характер возникновения уязвимостей, предполагая, что оно аппроксимируется нормальным законом распределения с параметрами $T_{\text{изм}}$ и $\sigma_{\text{изм}}$ (п. 4.2 исходных данных таблицы 1). Воспользовавшись правилом трех сигм [8], согласно которому значение случайной величины периодичности возникновения уязвимостей $t_{\text{изм}}$ практически не превышает ее отклонения от среднего значения $T_{\text{изм}}$ на величину $3\sigma_{\text{изм}}$, получим, что при выполнении условия (блок 8)

$$t_{\text{контр}} \geq t_{\text{изм}} = T_{\text{изм}} + 3\sigma_{\text{изм}} \qquad (9)$$

реализации угроз осуществляется чаще, чем расчетный по (6) контроль $t_{\text{контр}}$ и выполнение контроля необходимо.

Если (9) не выполняется, то к рассчитанному согласно (4) моменту времени $t_0 = t_{\text{контр}}$ какие-либо новые угрозы КП не предвидятся. Чтобы не допустить излишнего контроля при отсутствии угроз вычисляется вероятность того, что к моменту времени t_0 изменения были внесены (блок 12).

$$P_{\text{изм}}(t_{\text{изм}[-1]} < t < t_0) = F(t_0) - F\left(t_{\text{изм}[-1]}\right) \qquad (10)$$

$$P_{\text{угр}}(t_{\text{угр}} > t_0) = 1 - P(t_{\text{угр}} < t_0) = 1 - P_{\text{КП}}(t_0). \qquad (11)$$

Вероятность наступления двух независимых событий равна произведению вероятностей каждого из них. Так, если вероятность наступления событий (8) и (9) меньше допустимой вероятности эксплуатации уязвимости $P_{\text{ЭУ}} = 1 - P_{\text{треб}}$, т.е. условие (блок 14)

$$P_{\text{изм}} \cdot P_{\text{угр}} \leq 1 - P_{\text{треб}} \qquad (12)$$

не выполняется, то выполнение очередного контроля возможно отложить на время Δt (блок 15), которое определяется варьированием момента времени контроля t_0 до выполнения условия (12) (блок 14). По истечении времени Δt процедура расчета рациональных параметров контроля повторяется (блок 16).

В блоках 7 и 10 определяется вид контроля. На заключительном шаге частной методики расчета осуществляется вывод полученных результатов контроля (блок 11). Представленная частная методика расчета рациональных параметров контроля лежит в основе второго шага методики повышения защищенности сетей передачи данных.

На рис. 3 представлен график зависимости времени контроля ФКП СВТ от вероятности внесения несанкционированных изменений в КП. Данные для расчета времени контроля представлены в табл. 3.

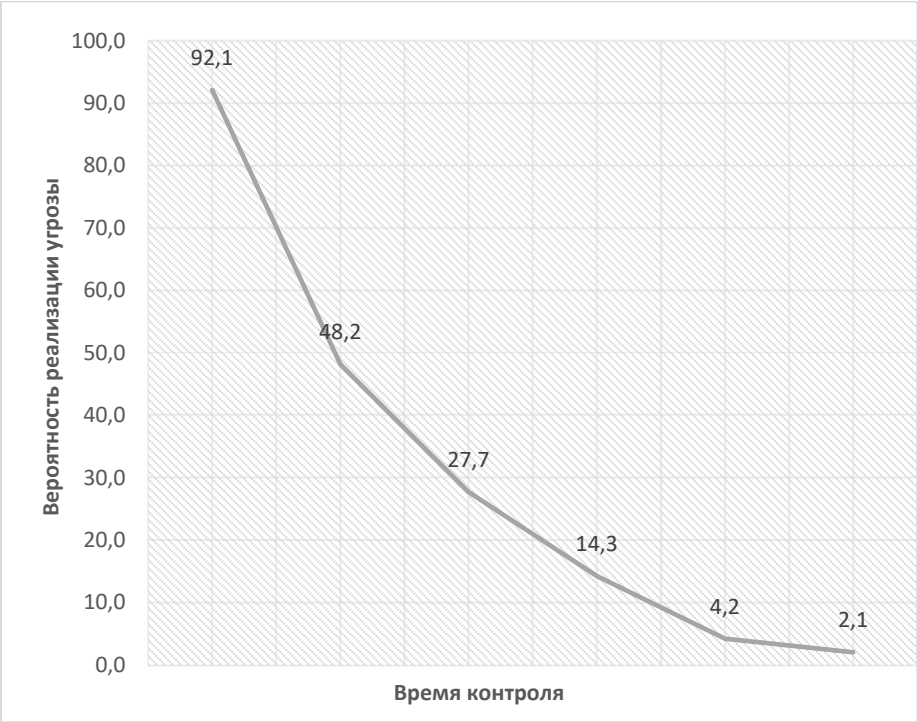

Рис. 3. Зависимость времени контроля
от вероятности реализации угрозы КП

Таблица 3

Экспериментальные данные
для расчета времени контроля неизменности ФК СВТ

Вероятность возникновения угроз целостности ФКП	Предполагаемое время до возникновения угрозы КП, мин	Время контроля неизменности ФКП, мин
0,1	40	92,1
0,3	40	48,2
0,5	40	27,7
0,7	40	14,3
0,9	40	4,2
0,95	40	2,1

В соответствии с представленным графиком, при одинаковом времени изменения КП, существенным для времени контроля, является вероятность изменения КП.

Для экономии ресурса контроля (вычислительного, временного, и материального) с применением полученных расчетных выражений, можно конфигурировать специальное программное обеспечение, предназначенное для контроля неизменности КП СВТ. В качестве основы, принята известная методика обеспечения целостности информации в программно-аппаратных комплексах связи за счет рационального резервирования [12].

Заключение

Рассмотрен подход к контролю неизменности контролируемых средств вычислительной техники. Разработанная методика в отличие от известных подходов к контролю предлагает обоснованные параметры контроля, рассчитанные на основе требуемой вероятности обеспечения целостности с учетом имеющихся ограничений.

Список литературы

1. Арсланов Х. А., Лихачев А. М. Актуальные научно-практические проблемы развития ОАЦСС ВС РФ // Связь в Вооруженных силах Российской Федерации – 2015. С. 29–36.

2. Макаренко С. И. Описательная модель сети связи специального назначения // Системы управления, связи и безопасности. 2017. № 2.

3. База данных уязвимостей ФСТЭК России. URL: https://bdu.fstec.ru/vul/

4. База данных угроз ФСТЭК России. URL: https://bdu.fstec.ru/threat/

5. Дождиков В. Г., Салтан М. И. Краткий энциклопедической словарь информационной безопасности. М. : ЭНЕРГИЯ, 2012. 240 с.

6. Лепешкин О. М., Шуравин А. С., Титов С. В. [и др.]. Методика контроля информационной безопасности распределенного узла связи // Известия Тульского государственного университета. Технические науки. 2020. № 12. С. 285–290.

7. Гмурман В. Е. Теория вероятностей и математическая статистика : учебник. М. : Юрайт, 2016. 479 с.

8. Рябинин И. А. Надежность и безопасность структурно-сложных систем. СПб. : Изд-во С.-Петерб. ун-та, 2007. 276 с.

9. Можаев А. С. Современное состояние и некоторые направления развития логико-вероятностных методов анализа систем. Ч. I. // Теория и информационная технология моделирования безопасности сложных систем : препринт / под ред. И. А. Рябинина. СПб. : ИПМАШ РАН, 1994. Вып. 1. С. 23–53.

10. Викторова В. С., Кунтшер Х., Петрухин Б. П., Степанянц А. С. Relex – программа анализа надежности, безопасности, рисков // Надежность. 2003. № 4 (7). С. 42–64.

11. Шуравин А. С., Новиков П. А. Современные угрозы безопасности информации, передаваемой с использованием узлов связи автоматизированной цифровой системы связи // Актуальные проблемы защиты и безопасности : тр. XXII Всерос. науч.-практ. конф. РАРАН. М., 2019. С. 181–183.

12. Киселев Д. В., Семенов С. С., Петров О. В. Методика обеспечения целостности информации в программно-аппаратных комплексах связи за счет рационального резервирования // Системы управления, связи и безопасности. 2019. № 1. С. 204–220. doi:10.24411/2410-9916-2019-10113

Для цитирования: Сычужников В. Б., Спирин С. В., Бакмаева К. Р., Евтихин И. О. Методика повышения защищенности сетей передачи данных на основе метода контроля уязвимостей // Безопасность информационных технологий : сб. науч. ст. по материалам IV Всерос. науч.-техн. конф. : в 2 т. Пенза : Изд-во ПГУ, 2022. Т. 1. С. 114–125.

УДК 621.37.39

МЕТОДИКА РАННЕГО ОБНАРУЖЕНИЯ КОМПЬЮТЕРНЫХ АТАК В СЕТЕВОМ ТРАФИКЕ СЕТИ ПЕРЕДАЧИ ДАННЫХ

О. С. Лаута[1], Э. А. Бударин[2], К. Р. Бакмаева[3], С. Н. Ракицкий[4]

[1,2,3,4] Военная академия связи имени Маршала Советского Союза С. М. Буденного Министерства обороны Российской Федерации, г. Санкт-Петербург

Аннотация. Рассмотрена методика выявления раннего обнаружения компьютерных атак с использованием гибридной нейронной сети. Представлена программная модель нейронной сети, которая позволяет выявлять ложные срабатывания. Приведен сравнительный анализ известных методов обнаружения компьютерных атак.

Ключевые слова: компьютерные атаки, сеть передачи данных, нейронная сеть, кибератаки, аномалии, фрактальный анализ

METHOD OF EARLY DETECTION OF COMPUTER ATTACKS IN THE NETWORK TRAFFIC OF THE DATA TRANSMISSION NETWORK

O. S. Lauta[1], E. A. Budarin[2], K. R. Bakmayeva[3], S. N. Rakitsky[4]

[1,2,3,4] Military Academy of Communications named after Marshal of the Soviet Union S. M. Budyonny of the Ministry of Defense of the Russian Federation, St. Petersburg

Abstract. The article discusses the method of detecting early detection of computer attacks using a hybrid neural network. A software model of a neural network is presented, which allows detecting false positives. A comparative analysis of known methods of detecting computer attacks is given.

Keywords: computer attacks, data transmission network, neural network, cyberattacks, anomalies, fractal analysis

Введение

Использование в сети передачи данных (СПД) информационных и коммуникационных технологий для сбора информации позволяет злоумышленнику воздействовать на сети путем компьютерных атак (КА). Этому способствует массовое применение устаревших операционных систем, малоэффективных механизмов защиты и наличие множественных уязвимостей в незащищенных сетевых протоколах. Подобные уязвимости помогают потенциальному злоумышленнику изменять настройки сетевых устройств, прослушивать

© Лаута О. С., Бударин Э. А., Бакмаева К. Р., Ракицкий С. Н., 2022

перенаправлять трафик, блокировать сетевое взаимодействие и получать несанкционированный доступ к внутренним компонентам СПД [1]. Воздействия КА приводит к появлению в СПД аномальной активности трафика. Все это послужило стимулом к поиску новых методов обнаружения и прогнозирования КА [2].

Основная часть

С целью постоянного мониторинга и обнаружения аномальной активности трафика в сети передачи данных (СПД) при компьютерных атаках (КА), их классификации, а также выявления ложных изменений в трафике, необходимо учитывать наличие большого количества сетевых маршрутов, на которых периодически возникают резкие колебания задержки в передаче данных и большие потери пакетов, появление новых свойств сетевого трафика, а также необходимость обеспечения высокого качества обслуживания приложений. Все это послужило поводом для разработки методики раннего обнаружения компьютерных атак в сетевом трафике СПД (рис. 1).

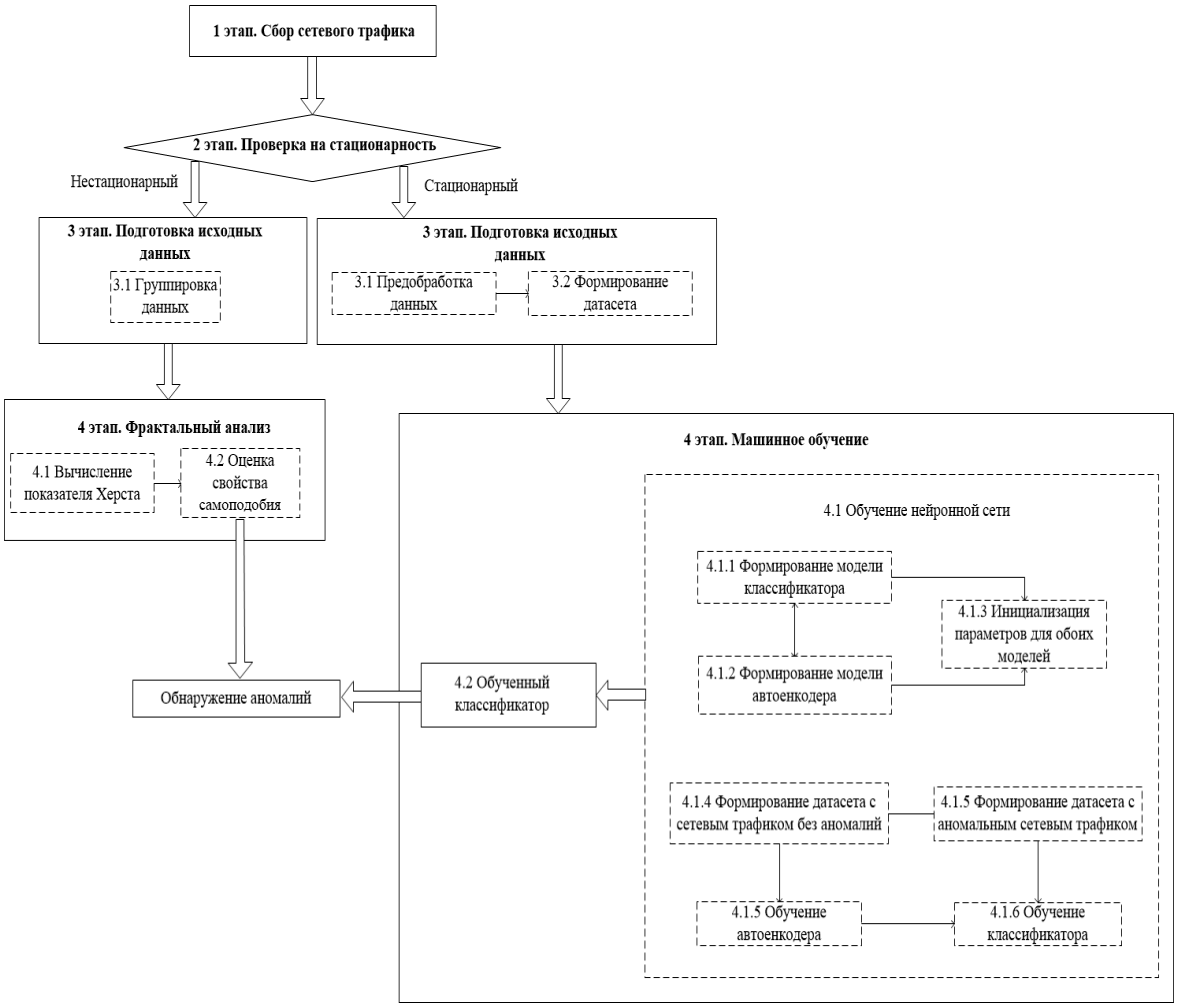

Рис. 1. Основные этапы методики раннего обнаружения компьютерных атак в сетевом трафике СПД

В качестве метода машинного обучения предлагается использовать гибридную нейронную сеть (рис. 2).

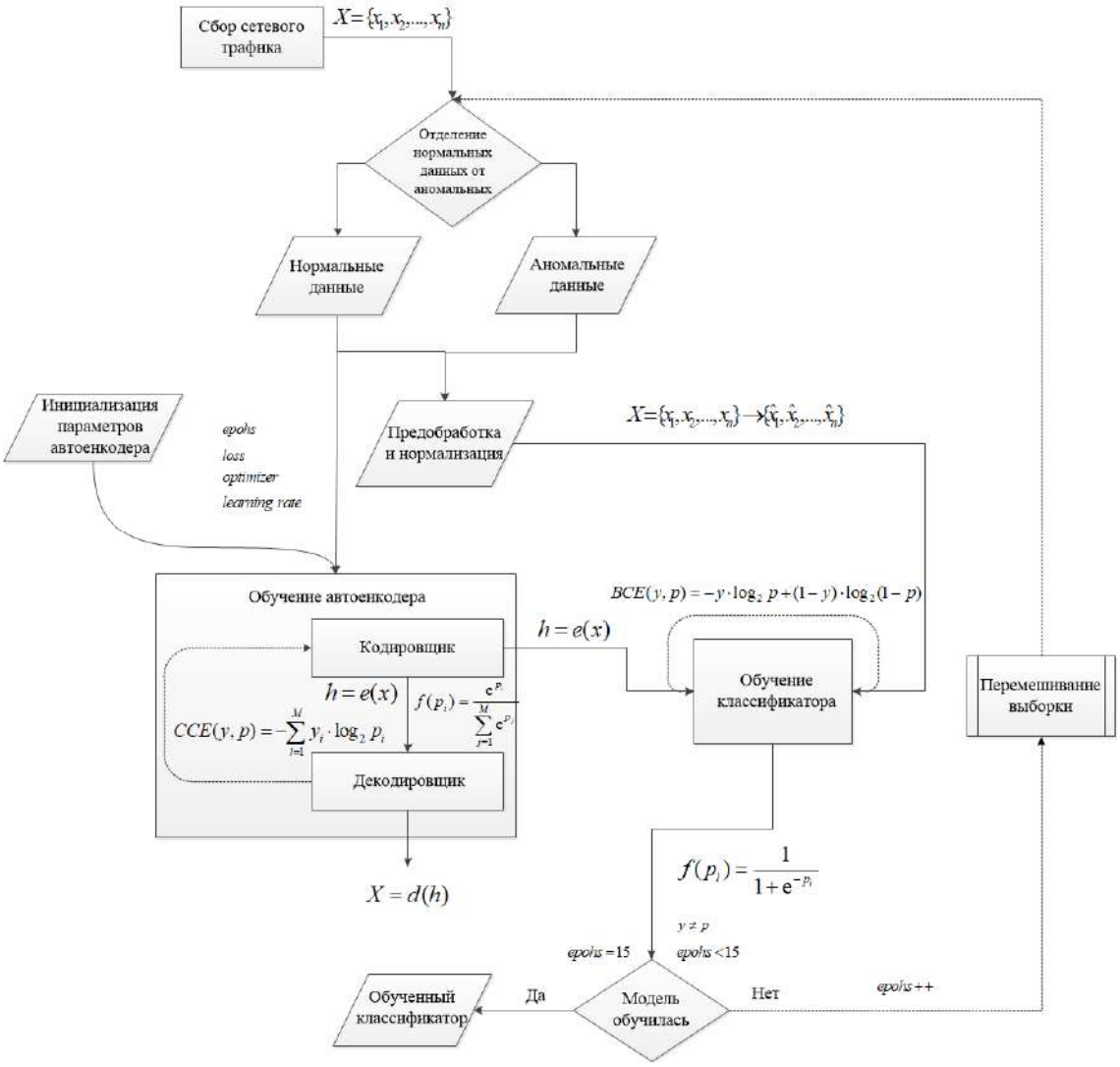

Рис. 2. Гибридная нейронная сеть

Для вычисления выходного значения сигналов нейронной сети, используется функция активации.

Активационной называется функция, аргументом которой является взвешенная сумма входов искусственного нейрона (1), а значением – выход нейрона (1):

$$p = \sum_{i=1}^{N} w_i x_i - b; \quad (1)$$

$$y = f(p), \quad (2)$$

где p – взвешенная сумма входов нейрона; N – число входов нейрона; x_i – входной сигнал; w_i – вес входного нейрона; b – смещение

выходного значения; $f(p)$ – активационная функция; y – выходное значение нейрона.

Для задачи бинарной классификации, лучше всего подходит логистическая (сигмоидальная) функция активации:

$$f(p_i) = \frac{1}{1+e^{-p_i}}, \tag{3}$$

$$p_i = \sum_{j=1}^{N} w_j x_j - b. \tag{4}$$

Эта функция относится к классу непрерывных функций и принимает на входе произвольное вещественное число, а на выходе выдает вещественное число в интервале от 0 до 1. В частности, большие (по модулю) отрицательные числа превращаются в ноль, а большие положительные – в единицу.

Функция потерь, которая используется для двоичной классификации, называется двоичной перекрестной энтропией (BCE). Эта функция эффективно наказывает нейронную сеть за ошибки двоичной классификации.

$$\text{BCE}(y, p) = -y \cdot \log_2 p + (1-y) \cdot \log_2 (1-p), \tag{5}$$

где y – двоичное истинное значение класса; p – предсказанное значение класса.

В автокодировщике используется функция softmax, которая применяется на выходном слое нейронной сети, для решения задач множественной классификации и представляет собой логистическую функцию, обобщенную для многомерного случая:

$$f(p_i) = \frac{e^{p_i}}{\sum_{j=1}^{M} e^{p_j}}, \tag{6}$$

где M – количество классов.

В таком случае в качестве функции потерь используется категориальная кросс энтропия, которая, в случае мульти-классовой классификации ($M > 2$), рассчитывается, как сумма значений логарифмических функций потерь для каждого прогноза наблюдаемых классов:

$$CCE(y, p) = -\sum_{i=1}^{M} y_i \cdot \log_2 p_i, \tag{7}$$

где y – истинное значение класса; p – предсказанное значение класса.

Обучение нейронной сети, осуществляется на основе алгоритма «обратного распространения», который позволяет рассчитать значения весовых коэффициентов таким образом, чтобы ошибка сети была минимальна. Для минимизации ошибки необходимо изменять веса в направлении противоположном градиенту. Соответственно для классификатора расчет весов будет производиться по формуле (8), а для автокодировщика (9):

$$\Delta w = -\alpha \cdot \frac{d\mathrm{BCE}}{dw}, \tag{8}$$

$$\Delta w = -\alpha \cdot \frac{d\mathrm{CCE}}{dw}, \tag{9}$$

где CCE, BCE – функции ошибок; Δw – величина, на которую необходимо изменить значение w; α – скорость обучения (learning rate).

Таким образом, для минимизации ошибки необходимо изменять w в направлении противоположном градиенту.

Программная модель нейронной сети

Модель нейронной сети состоит из рекуррентных ячеек – LSTM и GRU. На вход нейронной сети подаются данные размерностью до 699 символов. Выходных слоев у нейронной сети несколько. Выходной слой у автоенкодера имеет точно такую же размерность, как и входной. Выходной слой классификатора 1. Он определяет, является ли запрос аномальным или легитимным.

В качестве слоев автокодировщика, используются рекуррентные ячейки с долгой краткосрочной памятью – LSTM и GRU.

Свойство рекуррентности позволяет искусственной нейронной сети «обращаться» к результатам своей работы в прошлом, делать анализ предикций. Тем самым контекст решений в будущем будет зависеть не только от первичного глубокого обучения LSTM, но и её дальнейшей работы в потоке.

Сети LSTM являются подтипом более общих рекуррентных нейронных сетей (RNN). Ключевым атрибутом повторяющихся нейронных сетей является их способность сохранять информацию или состояние ячейки для дальнейшего использования в сети. Это делает их особенно подходящими для анализа временных данных, которые меняются с течением времени. Сети LSTM используются в таких задачах, как распознавание речи, перевод текста и, в данном случае, для обнаружения аномалий сети.

LSTM может удалять информацию из состояния ячейки; этот процесс регулируется структурами, называемыми фильтрами (gates).

Фильтры позволяют пропускать информацию на основании некоторых условий. Они состоят из слоя сигмоидальной нейронной сети и операции поточечного умножения. Сигмоидальный слой возвращает числа от нуля до единицы, которые обозначают, какую долю каждого блока информации следует пропустить дальше по сети. Ноль в данном случае означает «не пропускать ничего», единица – «пропустить все».

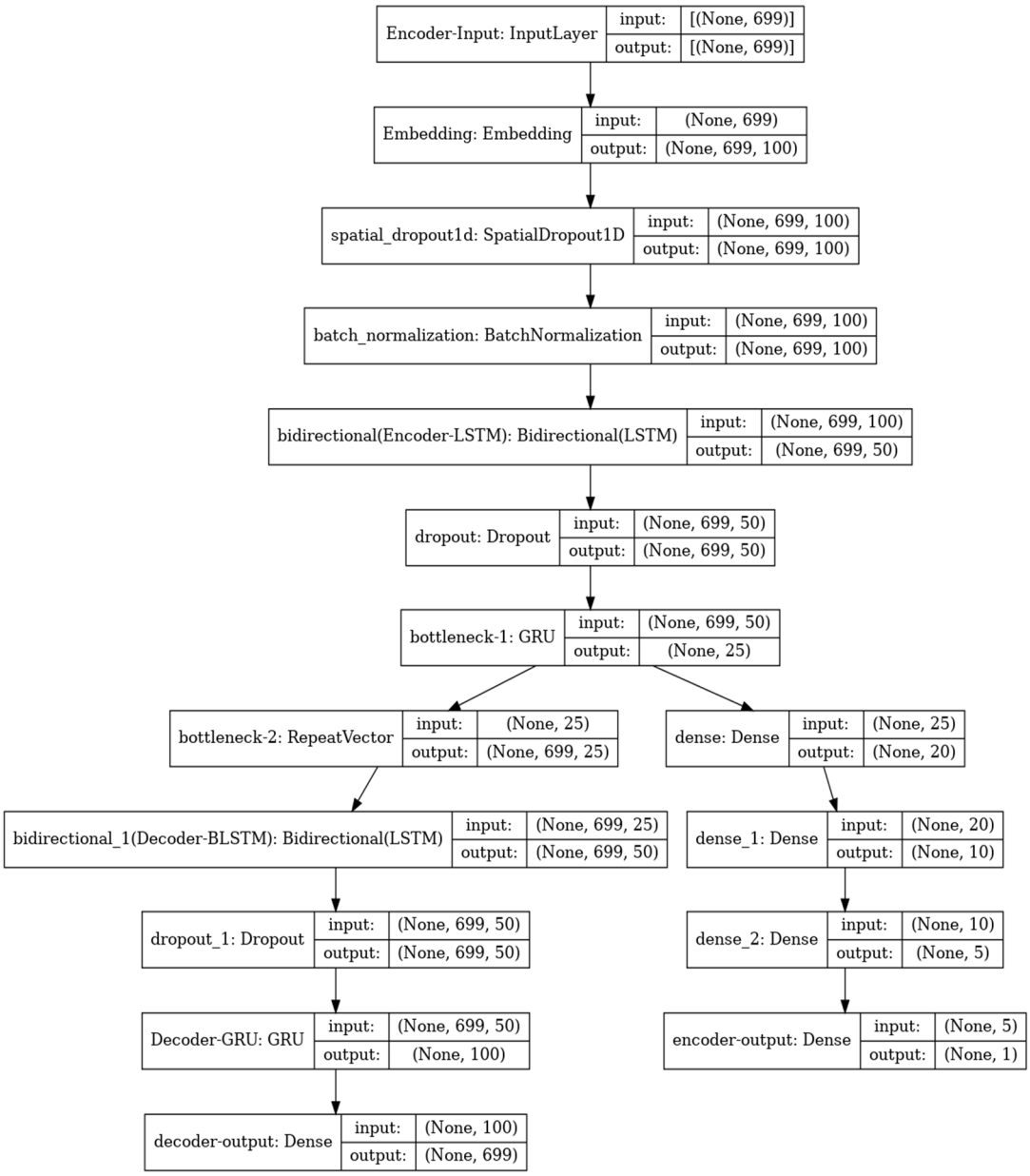

Рис. 3. Граф гибридной нейронной сети

Экспериментальная часть

Для формирования датасета, который годился бы для обучения предложенной архитектуры нейронной сети и оценки ее возможности

обнаруживать аномалии, необходимо было сначала смоделировать КА обхода WAF.

Общая идея нахождения способов обхода WAF заключается в приведении запроса к виду, в котором он остается понятным для атакуемого веб-приложения, но при этом является не понятным или кажется безобидным для WAF. Для этой цели могут использоваться различные спецсимволы, которые могут нарушать логику работы WAF и при этом быть понятными серверу.

На рис. 4 представлен пример обычного запроса и ответа на него сервера. Запрос представляет собой попытку передать методом POST параметр «select». Перехват запросов осуществлялся с помощью программы Burp Suite, предназначенной для выполнения тестов по безопасности веб-приложений.

Рис. 4. Попытка отправить запрос без преобразования параметра

Как видно из рис. 4 параметр «select» заблокирован WAF. На рис. 5 представлен другой запрос, в котором параметр «select» закодирован спецсимволами, взятыми из кодировочной таблицы Unicode.

Рис. 5. Попытка отправить запрос с кодированием параметра

Как видно из рис. 5, если передать параметр «select» в кодировке Unicode, то WAF его не обнаруживает.

Передадим теперь на сервер таким же способом (то есть с использованием кодировки Unicode) параметр «union select sleep(10)#» (рис. 6). Запрос также успешно обошел WAF и поступил на обработку на сервер.

Рис. 6. Пример обхода WAF зашифрованным запросом

Запрос, который был направлен на обработку на сервер, выглядит следующим образом:

SELECT note FROM notes WHERE assignee = «union select sleep(10)#».

Сервер вернул ответ на запрос с задержкой в 10 секунд. Тем самым была реализована эффективная кибератака вида «time base sql injection». Пример реализации этой кибератаки, осуществляющей обход WAF, представлен на рис. 7.

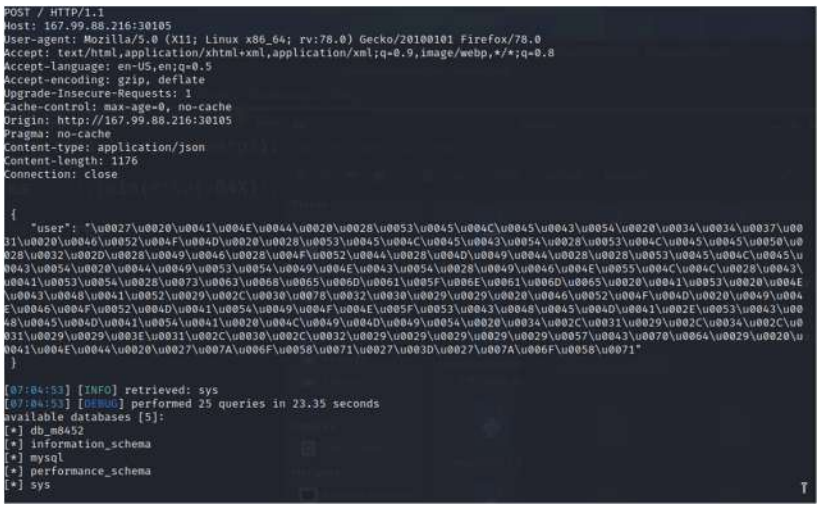

Рис. 7. Пример кибератаки обхода WAF

Существуют интеллектуальные WAF, основанные на методах машинного обучения, однако у них агрессивный режим работы и большое количество ложных срабатываний [3].

Программная реализация методики раннего обнаружения компьютерных атак в сетевом трафике сетей передачи данных

разработана на языке программирования Python с использованием библиотеки Pandas, с помощью которой осуществлялись обработки и анализа данных. Библиотека Pandas написана на языках программирования Си, Cython, and Python. Она делает Python мощным инструментом для анализа данных и дает возможность на высоком уровне строить сводные таблицы, выполнять группировки, предоставлять удобный доступ к табличным данным.

Кроме того, кроме библиотеки Pandas использовалась библиотека NumPy, которая представляет собой инструмент более низкого уровня, обеспечивающий работу с высокоуровневыми математическими функциями, а также с многомерными массивами (тензорами).

Графики строились с помощью модуля Matplotlib на основе полученного набора данных. Все расчеты производились в интегрированной среде разработки Jupiter notebook.

На рис. 8 изображена блок схема, отражающая этапы формирования данных и обучения гибридной модели нейронной сети.

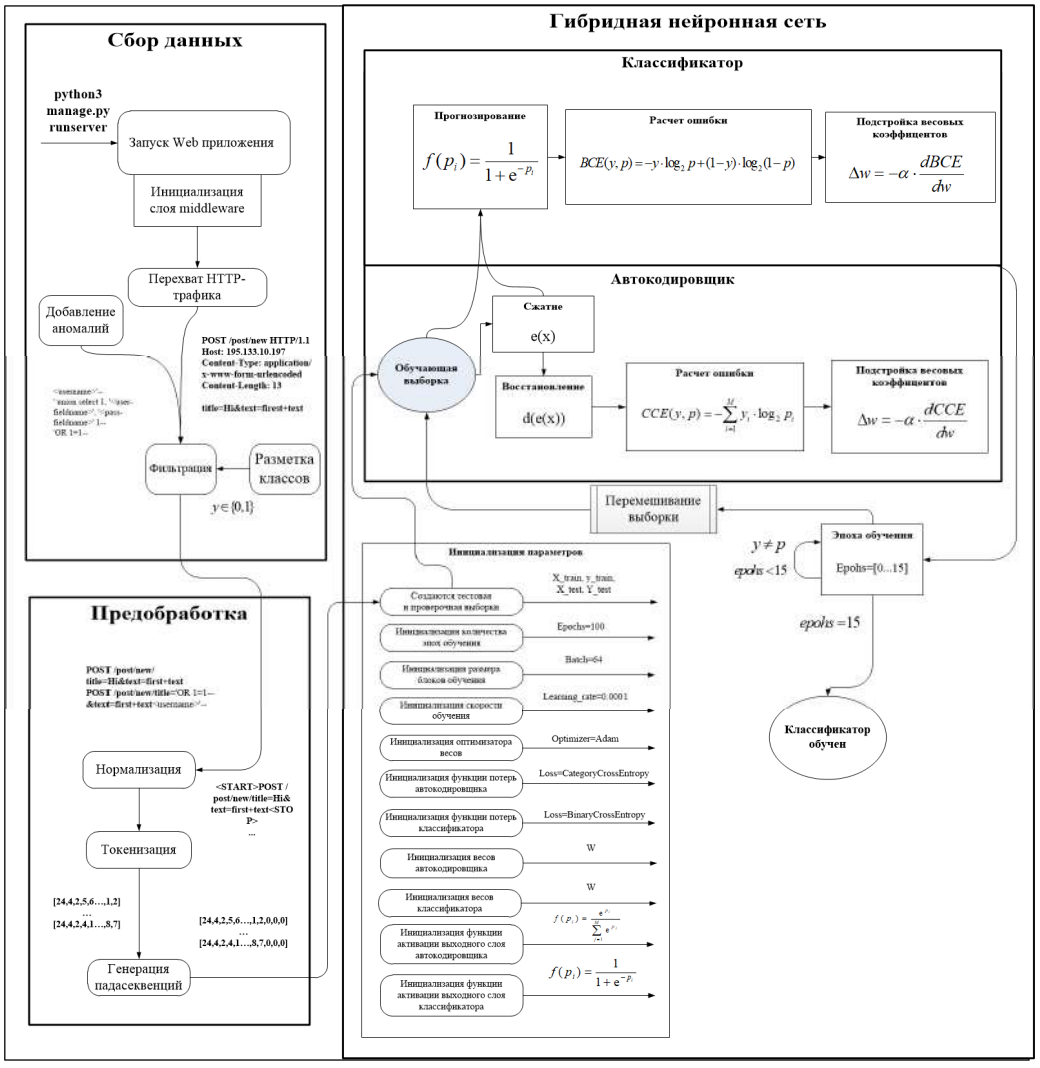

Рис. 8. Блок-схема этапов обучения гибридной нейронной сети

Для того, чтобы сформировать датасет данных, предназначенных для обучения нейронной сети, на языке программирования python реализовано web-приложение способное перехватывать любые пользовательские запросы с помощью промежуточного слоя middleware. Такой подход позволяет обрабатывать запросы из браузера, прежде чем они, достигнут представления Django (сервера), а также ответы от представлений до того, как они возвращаются в браузер.

Следующий этап обучения гибридной нейронной сети включает в себя нормализацию данных. Запросы оборачиваются специальными токенами <START> и <STOP>, что задаёт верное вероятностное распределение над последовательностями разной длины [4].

В книге [5] изучалось распределение расстояний (т.е. количество символов) между одинаковыми буквами в тексте. Оказалось, что это квазистационарный ряд, одинаково распределенный (по гамма-распределению) для любой буквы алфавита, и близкий к белому шуму. Исходя из этого, для проверки стационарности HTTP-трафика, проведен эксперимент, который заключался в построении графика распределения длин между двумя одинаковыми символами (рис. 9) и оценки стационарности получившегося ряда с помощью теста Дики-Фуллера.

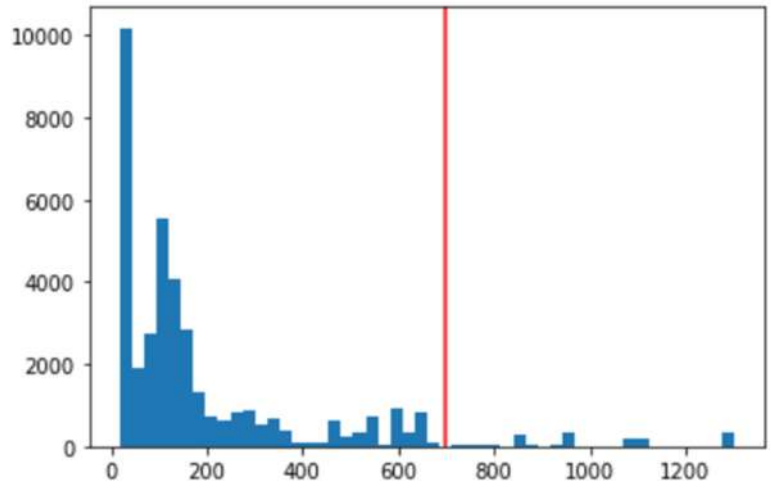

Рис. 9. Распределение длин легитимных запросов

Далее произведена предобработка и нормализация получившейся выборки. Поскольку протокол HTTP – текстовый протокол, использовалось векторное представление символов. Для этого сперва осуществляется замена символов, встречающихся в датасете на числовой эквивалент, который не имеет самостоятельного смысла/значения для внешнего или внутреннего использования (токенизировать),

а затем переводятся слова в последовательность секвенций (рис. 10), с помощью токенизации.

```
data_train = pad_sequences(X_train_sequences, maxlen=max_len_str, padding='post')
data_test = pad_sequences(X_test_sequences, maxlen=max_len_str, padding='post')
```

```
data_train
```

```
array([[24,  4,  2, ...,  0,  0,  0],
       [24,  4,  2, ...,  0,  0,  0],
       [24,  4,  2, ...,  0,  0,  0],
       ...,
       [24,  4,  2, ...,  0,  0,  0],
       [24,  4,  2, ...,  0,  0,  0],
       [24,  4,  2, ...,  0,  0,  0]], dtype=int32)
```

Рис. 10. Секвенции

При этом, важным обстоятельством является то, что все секвенции должны быть одной длины. Если запрос меньше длины секвенции, то оставшиеся символы заполняются нулями.

Далее секвенции подаются на вход гибридной нейронной сети, и подбираются гиперпараметры. Но перед этим, необходимо настроить среду выполнения таким образом, чтобы все вычисления происходили на GPU. Для этой цели достаточно установить библиотеки, строго определенных версий:

```
conda create --name tf python=3.8
conda activate tf
conda install cudatoolkit=10.0.130
conda install cudnn=7.6.0=cuda10.0_0
pip install --upgrade tensorflow-gpu
sudo apt install libcudnn8
```

Датасет предназначенный для обучения нейронной сети должен включать как легитимный трафик, так и аномальный. На вход автокодировщика подается только легитимный трафик. На вход классификатора подается легитимный, аномальный трафик и скрытые латентные представления, полученные с автокодировщика после кодирования информации (рис. 11).

Подбор параметров осуществляется таким образом, чтобы функция потерь при обучении автокодировщика уменьшалась, при этом точность классификатора росла.

На рис. 12 продемонстрирован рост точности и снижение потерь на 30 эпохах обучения.

```
Trial 27 Complete [00h 27m 58s]
decoder-output_loss: 19800.93359375

Best decoder-output_loss So Far: 19800.93359375
Total elapsed time: 05h 00m 36s

Search: Running Trial #28

Hyperparameter    |Value             |Best Value So Far
decoder-output    |0.0014            |0.0039
encoder-output    |90                |45
learning_rate     |1e-06             |1e-05
tuner/epochs      |10                |10
tuner/initial_e...|0                 |0
tuner/bracket     |0                 |0
tuner/round       |0                 |0

Epoch 1/10
1351/1351 [==============================] - 218s 156ms/step - loss: 92.7079 - encoder-
Epoch 2/10
1351/1351 [==============================] - 208s 154ms/step - loss: 92.4347 - encoder-
Epoch 3/10
1026/1351 [=====================>........] - ETA: 43s - loss: 92.1161 - encoder-output_
```

Рис. 11. Подбор гиперпараметров нейронной сети

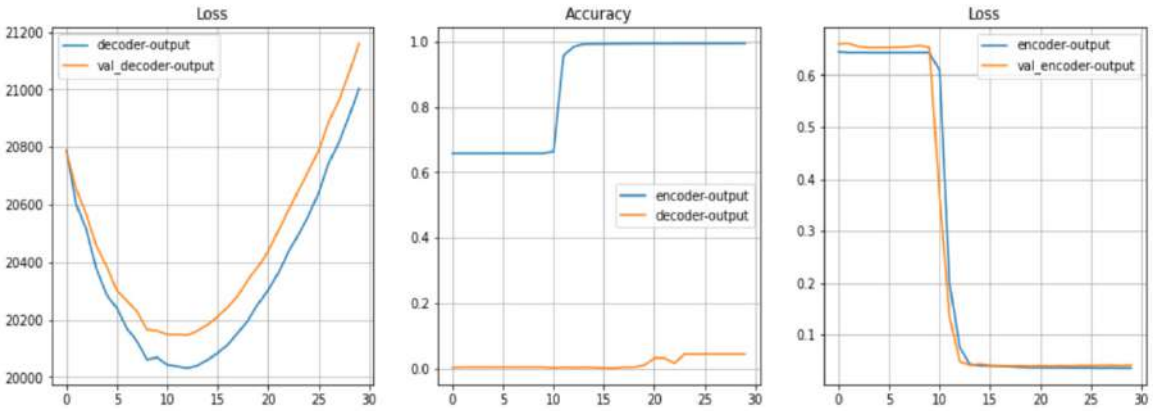

Рис. 12. Обучение декодера и классификатора на 30 эпохах

После обучения нейронной сети, проведен эксперимент по оценке точности и полноты. Сперва использовался датасет с КА такого же типа, как в датасете при обучении модели. Результат обнаружения аномалий – 96,9 % (рис. 13).

```
scores = encoder.evaluate(X_pad, data_last['anomaly'].values, verbose=1)
print('Точность: {}% \nLoss: {}'.format(scores[1]*100, 1 - scores[1]))

1799/1799 [==============================] - 52s 29ms/step - loss: 0.0478 - binary_accuracy: 0.9850
Точность: 96.90268635749817%
Loss: 0.03097313642501831
```

Рис. 13. Оценка точности алгоритма на известных аномалиях

Затем был сформирован новый датасет, с компьютерными атаками раннее неизвестными классификатору (атаки нулевого дня) и алгоритм распознал 99 % неизвестных раннее атак (атаки 0-го дня),

и верно определил, что 99 % легитимных запросов не являются аномальными (рис. 14).

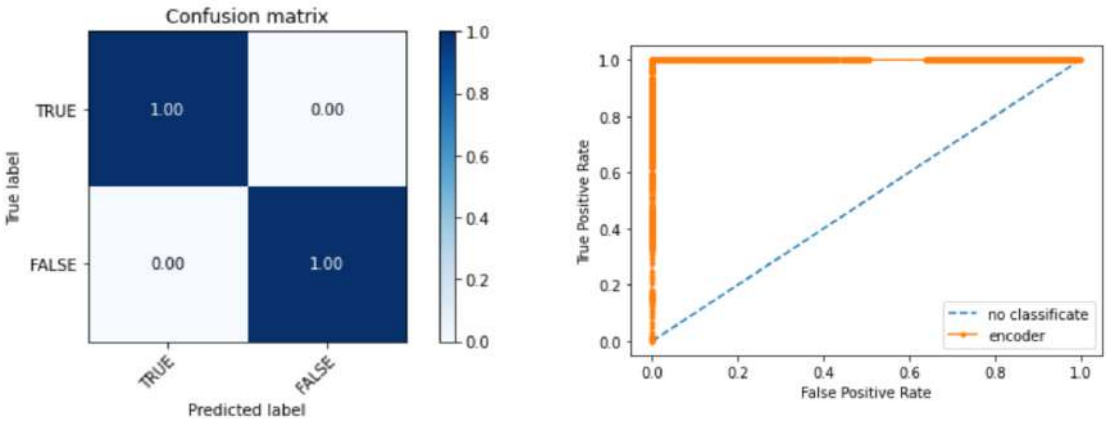

Рис. 14. Оценка точности алгоритма на неизвестных аномалиях

Система допускает ложные срабатывания. В данном случае, 10 запросов были отброшены нейронной сетью (рис. 15).

```
data_last[(data_last['pred']==True) & (data_last['anomaly']==False)]
```

	data	anomaly	pred
661	<START>POST /post/new/title=Действительно,+бол...	0	True
1153	<START>POST /post/new/title=англоязычный+форум...	0	True
2191	<START>POST /post/new/title=Ведь+в+понятиетext...	0	True
2662	<START>POST /post/new/title=ладно,как+начну+сн...	0	True
5775	<START>POST /post/new/title=The+cyber-physical...	0	True
6800	<START>POST /post/new/title=PopMech+and+its+an...	0	True
11024	<START>POST /post/new/title=ладно,как+начну+сн...	0	True
22111	<START>POST /post/new/title=Google+Search+Rank...	0	True
27957	<START>POST /post/new/title=огромное+спасибо+з...	0	True
34286	<START>POST /post/new/title=BlueWillowPlate++t...	0	True

Рис. 15. Пример ложных срабатываний

Учитывая тот факт, что в DataFrame содержится 57.000 запросов, из которых 20.000 аномальных, числом 10 не является существенным недостатком текущего подхода.

Сравнительная оценка эффективности рассматриваемой системы была проведена на основе его сравнения с другими известными системами, например, IDS и IPS, которые при обнаружениях компьютерных атак (аномалий) используют сигнатурный метод, статисти-

ческий метод и методы машинного обучения. Результаты такого сравнения представлены в табл. 1. Для сравнения использовалась трёхбалльная шкала: «высокая», «средняя», и «низкая».

В качестве основных учитываемых параметров сравниваемых методов рассматривались скорость и точность обнаружения кибератак, как известных, так и неизвестных, возможность работы со стационарным и нестационарным трафиком, а также то, насколько часто происходят ложные срабатывания.

Таблица 1

Результаты сравнительного анализа известных методов обнаружения кибератак

Наименование метода	Скорость обнаружения	Точность обнаружения		Тип трафика		Отсутствие ложного обнаружения
		Для известных атак	Для неизвестных атак	Стационарный	Нестационарный	
Сигнатурные методы	высокая	высокая	низкая	высокая	низкая	средняя
Статистические методы	средняя	средняя	средняя	высокая	низкая	средняя
Методы машинного обучения	средняя	высокая	средняя	высокая	низкая	высокая
Предлагаемый метод (фрактальный анализ и нейронные сети)	высокая	высокая	средняя	высокая	высокая	средняя

Сигнатурные методы используют заранее составленные правила. Поэтому они достаточно быстрые и имеют высокую точность обнаружения известных типов кибератак. Однако они не способны обнаруживать новые, неизвестные типы атак, включая таргетированные атаки. Кроме того, они имеют средние показатели по отсутствию ложного обнаружения.

Статистические методы используют накопленную статистику. По этой причине они уступают сигнатурным методам по скорости обнаружения и точности обнаружения известных атак. В то же время в ряде случаев они способны обнаруживать неизвестные атаки. По ложному обнаружению они имеют примерно такие же возможности, как и сигнатурные методы.

Методы машинного обучения в настоящее время являются достаточно разнообразными и хорошо развитыми. Учитывая, что в этих методах процессу обнаружения атак обязательно предшествует

процесс обучения на контрольной выборке, мы считаем, что по скорости обнаружения атак эти методы уступают сигнатурным методам. Однако эти методы имеют более высокую точность обнаружения известных атак и хорошую точность обнаружения неизвестных атак. При этом доля ложных срабатываний в методах машинного обучения является крайне низкой.

В тоже время как показали эксперименты, сетевой трафик СПД обладает фрактальными свойствами. Иными словами, на больших объемах этот трафик обладает свойством самоподобия.

Кроме того, эксперименты продемонстрировали, что предлагаемая проактивная система защиты СПД при обнаружении КА на основе оценки самоподобия параметров функционирования системы с использованием фрактальных показателей и прогнозирования факта воздействия кибератак путем применения предложенной структуры нейронной сети LSTM обладает достаточно высокой эффективностью при обнаружении как известных, так и неизвестных КА. Вероятность обнаружения известных КА равна 0,96, а атаки «нулевого дня» – 0,8.

Главным достоинством предлагаемого подхода является высокая скорость детектирования аномалий, вызываемых КА, возможность работы с любыми видами трафика, а также низкая вероятность ложного срабатывания. Определение коэффициента Херста производится за несколько микросекунд в зависимости от производительности вычислительной техники, а классификация аномалий с помощью нейронной сети LSTM позволяет максимальность снизить вероятность ложного срабатывания.

В то же время, эксперименты показали, что наиболее популярные алгоритмы машинного обучения, протестированные на сгенерированных временных рядах, отлично справляются с обнаружением аномальных выбросов. В таком случае, аномалия проявляется в виде нестационарности некоторых наблюдаемых временных рядов. Это не только мгновенные скачки амплитуды измерений, но и медленные тренды, практически невидимые за время наблюдений.

Заключение

В статье разработан новый подход к работе системы защиты СПД от КА, которая основана на применении основных положений теории фракталов и использовании предлагаемых этой теорией методов оценки самоподобия, таких как тест Дики-Фуллера, R/S анализ и метод DFA. При тестировании фрактальных методов, позволяющих проводить исследования долговременных зависимостей в трафике

компьютерной сети, метод DFA является более эффективным, чем R/S анализ из-за его возможности обрабатывать не только стационарные, но и нестационарные ряды с высокой точностью. Совместное его применение с LSTM сетями позволяет существенно увеличить вероятность обнаружения КА.

Специфика предложенной системы является то, что обнаружение КА производится с использованием автокодировщика, обученного на основе эталонных данных работы СПД и информационного обмена в ней, учитывающего все отклонения от штатной работы сети. В процессе работы автокодировщик дополнительно обучается обоснованной нейронной сетью, т.е. в итоге получается генеративно-состязательная сеть, в которой нейронные сети учатся друг у друга.

Проведенная экспериментальная оценка предложенного подхода показала, что, по сравнению со многими другими подходами, одним из главных преимуществ фрактального анализа является скорость его работы, а также возможность обнаружения аномалий в трафике любого вида. К увеличению времени расчета приводит только увеличение количества обрабатываемых параметров заголовка протокола передачи данных (длина пакета, флаги и т.д.). При этом предложенная система продемонстрировала достаточно высокую вероятность обнаружения КА, достигнув значения 0,96 для известных атак и 0,8 для ранее неизвестных атак.

Предложенная система может являться элементом используемых в настоящее время IDS и IPS, основной задачей которых является анализ передаваемых внутренние потоки данных, находя в них последовательности битов, которые могут представлять из себя вредоносные действия или события, а также осуществляют мониторинг системных журналов и других файлов регистрации деятельности пользователей. Она позволит повысить вероятность обнаружения неизвестных КА за счет архитектуры автокодировщика, уменьшить вероятность ложного срабатывания, время и объем оперативной памяти, задействованных для анализа сетевого трафика, что нивелирует недостатки существующих IDS и IPS, основанных на жестких правилах, а также сигнатурной и аномальной технологиях.

Кроме того, следует отметить, что проведенные исследования демонстрируют пока еще только потенциальную возможность и эффективность предложенной системы к прогнозированию и обнаружению КА в сети СПД. Практическая реализация и дальнейшее совершенствование этой системы, её распространение на различные виды КА, а также ее взаимодействие с другими методами являются дальнейшими направлениями исследований.

Список литературы

1. Kotenko I., Saenko I., Lauta O., Kribel A. An approach to detecting cyber attacks against smart power grids based on the analysis of network traffic self-similarity // Energies. 2020. Т. 13, № 19. P. 5031.

2. Крибель А. М., Лаута О. С., Филин А. В., Фень А. С. Метод обнаружения аномалий в сетевом компьютерном трафике на основе нейронной сети с использованием LSTM // Электросвязь. 2021. № 12. С. 43–48.

3. Карпов М. А., Коцыняк М. А., Нечепуренко А. П. Модель функционирования информационно-телекоммуникационной сети специального назначения в условиях информационного воздействия // Актуальные проблемы защиты и безопасности : тр. XXIV Всерос. науч.-практ. конф. РАРАН : в 7 т. (г. Санкт-Петербург, 31 марта – 3 апреля 2021 г.). М. : Российская академия ракетных и артиллерийских наук, 2021. Т. 1. С. 458–462.

4. Орлов Ю. Н., Осминин К. П. Методы статистического анализа литературных текстов. М. : Editorial URSS : ЛИБРОКОМ, 2012. 312 с.

5. Bale J. P. M., Sediyono E. & oth. Facilitated Risk Analysis Process. Risk management in information technology using facilitated risk analysis process (FRAP). Academic information systems of Satya Wacana Christian University, 2015.

Для цитирования. Лаута О. С., Бударин Э. А., Бакмаева К. Р., Ракицкий С. Н. Методика раннего обнаружения компьютерных атак в сетевом трафике сети передачи данных // Безопасность информационных технологий : сб. науч. ст. по материалам IV Всерос. науч.-техн. конф. : в 2 т. Пенза : Изд-во ПГУ, 2022. Т. 1. С. 126–142.

УДК 621.37.39

МОДЕЛЬ ВЫЯВЛЕНИЯ АНОМАЛИЙ В СЕТЕВОМ ТРАФИКЕ СЕТИ ПЕРЕДАЧИ ДАННЫХ В УСЛОВИЯХ КОМПЬЮТЕРНЫХ АТАК

Р. А. Перов[1], С. Ю. Скоробогатов[2], Э. А. Бударин[3], В. Б. Сычужников[4]

[1,2,3,4] Военная академия связи имени Маршала Советского Союза С. М. Буденного Министерства обороны Российской Федерации, г. Санкт-Петербург

Аннотация. Предложена модель выявления аномалий, вызванных воздействием компьютерных атак в сети передачи данных, позволяющая в реальном или близком к реальному масштабу времени их выявлять. Экспериментальные результаты также свидетельствуют о том, что при появлении сетевых аномалий, вызванных, например, кибератаками типа DDoS и «сканирование сети и ее уязвимостей», характер этих свойств начинает существенно отличаться от нормального трафика.

Ключевые слова: аномалии, компьютерные атаки, сеть передачи данных, показатель Херста, самоподобие, временной ряд

A MODEL FOR DETECTING ANOMALIES IN THE NETWORK TRAFFIC OF A DATA TRANSMISSION NETWORK IN THE CONDITIONS OF COMPUTER ATTACKS

R. A. Perov[1], S. Y. Skorobogatov[2], E. A. Budarin[3], V. B. Sychuzhnikov[4]

[1,2,3,4] Military Academy of Communications named after Marshal of the Soviet Union S. M. Budyonny of the Ministry of Defense of the Russian Federation, St. Petersburg

Abstract. The article proposes a model for detecting anomalies caused by the impact of computer attacks on the data transmission network, which allows them to be detected in real or close to real time. Experimental results also indicate that when network anomalies appear, caused, for example, by cyber-attacks such as DDoS and "scanning the network and its vulnerabilities", the nature of these properties begins to differ significantly from normal traffic.

Keywords: anomalies, computer attacks, data transmission network, Hurst index, self-similarity, time series

Современный этап развития общества характеризуется повышением роли информационной сферы, представляющей собой совокуп-

© Перов Р. А., Скоробогатов С. Ю., Бударин Э. А., Сычужников В. Б., 2022

ность информации и информационных технологий, что позволило осуществлять сбор, формирование, хранение, обработку и распространение информации в таких объемах и с такой оперативностью, которые были немыслимые раньше.

Именно новые технологии привели к бурному распространению сетей передачи данных (СПД), открывающих принципиально новые возможности международного информационного обмена. Происходит интеграция и конвергенция сетей и служб. Это обеспечивает доступ пользователей к любой услуге, имеющейся во множестве сетей, за счет гибких возможностей по их обработке и управлению.

Несмотря на удобство, экономическую выгоду и эффективность использования СПД, а также, темпы, с которыми развивается современная сфера информационных технологий, подвергают мировое сообщество целому ряду беспрецедентных угроз и факторов уязвимости, которые злоумышленнику открывают возможность реализации компьютерных атак (КА).

Структура модели сетевого трафика СПД

Для постоянного мониторинга и обнаружения аномальной активности трафика в СПД необходимо учитывать наличие большого количества сетевых маршрутов, на которых периодически возникают резкие колебания задержки в передаче данных и большие потери пакетов, появление новых свойств сетевого трафика, а также необходимость обеспечения высокого качества обслуживания приложений [1]. Именно поэтому на первоначальном этапе важно определиться с моделью, которая будет максимально точно описывать сетевой трафик.

Для создания адекватной модели требуется использование наиболее подходящего математического аппарата. Принимаемая для описания модель должна быть, по возможности, максимально близка к описываемому реальному процессу [2]. Оценить степень близости модели и реального процесса не всегда возможно, поскольку в некоторых случаях реальные процессы попросту недоступны для проведения и наблюдений. В таких случаях приходится полагаться на те логические и иные доводы, которые принимаются при выборе определенной модели и ее параметров.

Трафик как случайный процесс характеризуется параметрами, которые определяют его основные, наиболее важные для моделирования, свойства [3]. Основной задачей модели трафика является описание поступающего потока при помощи набора параметров таким образом, чтобы эти выбранные значения параметров можно было бы применить для нахождения аномалий и вредоносной активности в сети.

Существуют модели, которые описывают сетевой трафик с помощью методов теории вероятностей и математической статистики,

а также теории массового обслуживания [4]. Как правило, такие процессы обладают свойством стационарности – вероятностные характеристики (среднее значение и дисперсия) не меняются с течением времени.

Наиболее простой, часто используемой стационарной моделью является модель простейшего (стационарного пуассоновского) потока [5]. Основным свойством потока является то, что количество пакетов, поступающих за заданный интервал времени, случайная величина, которая подчиняется распределению Пуассона, а интервалы времени между пакетами случайны и подчиняются экспоненциальному распределению. Модель простейшего потока часто применяется для описания трафика, производимого большим количеством независимых источников, например, трафика в сетях с коммутацией каналов.

В сетях с коммутацией пакетов свойства потоков не всегда могут быть описаны распределением Пуассона [6], ввиду нестационарности потока.

Поэтому выделяют нестационарные модели, способные более корректно, описывать сетевой трафик для СПД с коммутацией пакетов. Такие модели основываются на фрактальном анализе и рассматривают сетевой трафик, как самоподобный нестационарный процесс. Под самоподобием понимается свойство сетевого трафика сохранять свой характер при изменении масштаба времени.

Впервые о самоподобном потоке заговорили еще в 1993 г. Leland, Taqqu, Willinger и Wilson проводили исследования Ethernet-трафика в сети корпорации Bellcore и пришли к выводу, что на больших интервалах он обладает свойством самоподобия, т.е. выглядит качественно одинаково при любых масштабах временной оси.

Самоподобие проявляется в том, что имеется медленно убывающая зависимость между величинами трафика в разные моменты времени, а число переданных пакетов имеет сходный вид в различных временных масштабах. Другими словами, самоподобные потоки зависят не только от времени, но и от предыдущих событий.

Постановка задачи исследования

При проектировании системы защиты, необходимо учесть все вышеперечисленное и разработать Модель выявления аномалий в сетевом трафике СПД в условиях КА, которая описывает сетевой трафик сразу двух видов: стационарный и нестационарный.

Исходные данные модели:

Hurst – показатель Херста; p – порядок авторегрессии (зависимость между наблюдениями и число интегрированных наблюдений); q – порядок скользящего среднего (зависимость между наблюдениями и остатками при применении модели к интегрированным

наблюдениям); $Z(t)$ – реальный сетевой трафик; Y – конечное множество меток класса (аномалия, не аномалия).

Назначение и цель:

Модель выявления аномалий в сетевом трафике СПД предназначена для описания и проверки сетевого трафика на стационарность. После проверки выбирается метод, который будет производить оценку сетевого трафика на наличие аномалий.

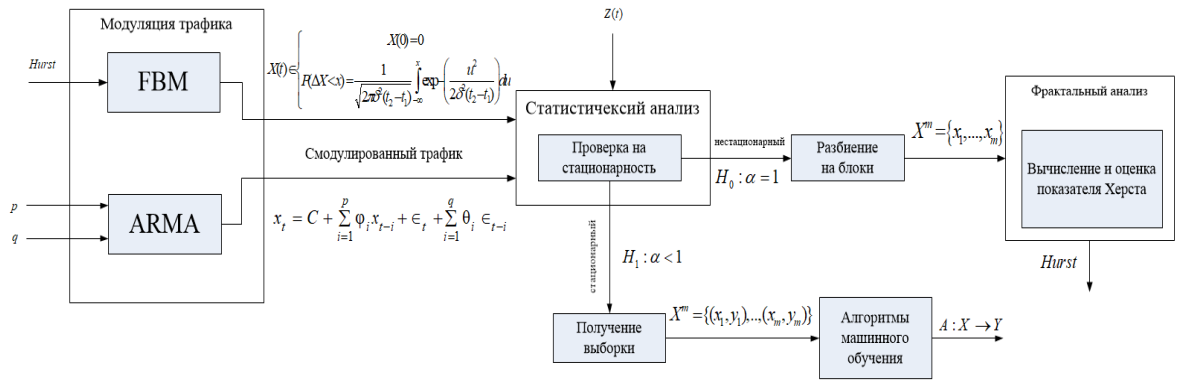

Рис. 1. Модель выявления аномалий в сетевом трафике СПД

Постановка задачи:

Требуется разработать модель, которая будет максимально точно описывать сетевой трафик в узлах СПД, учитывая, как случайный $X(t)$, так и стационарный (детерминированный) x_t процесс. Кроме того, модель должна производить проверку на стационарность $H_1 : \alpha < 1$ не только сгенерированного трафика, но и реального $Z(t)$, полученного в ходе эксперимента. Также модель должна принимать решение по выбору алгоритма, с помощью которого будет производиться оценка сетевого трафика на наличие аномальной активности. В случае нестационраности $H_0 : \alpha = 0$ сетевого трафика, оценка производиться на основании вычисления показателя *Hurst*. При $H_1 : \alpha < 1$ с помощью методов машинного обучения, находится целевая зависимость между аномалиями и признаками сетевых фреймов.

Выходные данные модели:

Hurst – показатель Херста;

$A : X \to Y$ – классификация объектов $x \in X$, где X множество сетевых фреймов.

Проверка на стационарность

Для проверки гипотезы о стационарности ряда используется расширенный тест Дики – Фуллера.

При помощи этого теста проверяют значение коэффициента авторегрессии α в авторегрессионном уравнении AR. Рассмотрим авторегрессионное уравнение первого порядка AR(1):

$$y_t = \alpha \cdot y_{t-1} + \varepsilon_t, \qquad (1)$$

где y_t – временной ряд, а ε – белый шум, $t = 1,...,T$.

1. Если $H_1 : \alpha < 1$, то ряд y_t будет стационарным, $y_t \sim I(0)$ и OLS-оценка $\hat{\alpha}$ будет иметь нормальное распределение с нулевым средним и дисперсией $1-\alpha^2$.

Для тестирования гипотезы единичного корня строится OLS-оценка $\hat{\alpha}$:

$$\hat{\alpha} = \frac{\sum_{t=1}^{T} y_{t-1} y_t}{\sum_{t=1}^{T} y_{t-1}^2}, \qquad (2)$$

И соответствующая ей t-статистика

$$t_\alpha = \frac{\hat{\alpha} - 1}{S / \sqrt{\sum_{t=1}^{T} y_{t-1}^2}}, \qquad (3)$$

где $S^2 = T^{-1} \sum_{t=1}^{T} (y_t - \hat{\alpha} y_{t-1})^2$ – оцененная дисперсия остатков.

Если $t_\alpha < t_{\text{табл}}^{5\%}$ – временной ряд стационарен на уровне значимости 5 %.

2. Если $H_0 : \alpha = 1$, то распределение этой оценки больше не будет нормальным, и процесс y_t будет нестационарным с зависящей от времени дисперсией $y_t \sim I(1)$. В этом случае для моделирования динамики такого ряда необходимо использовать его первую разность $\Delta y_t = y_y - y_{t-1}$. При нулевой гипотезе статистика нормализованного смещения $T(\hat{\alpha} - 1)$ и t-статистика t_α имеют нестандартные предельные распределения Дики – Фуллера:

$$T(\hat{\alpha} - 1) \Rightarrow \frac{\int_0^1 W(r)\, dW(r)}{\int_0^1 W^2(r)\, dr}, \qquad (4)$$

$$t_\alpha \Rightarrow \frac{\int_0^1 W(r)\, dW(r)}{\sqrt{\int_0^1 W^2(r)\, dr}}, \qquad (5)$$

где $W(r)$ – стандартный Винеровский процесс (Броуновское движение).

Если $t_{\alpha} > t_{\text{табл}}^{5\,\%}$ – временной ряд нестационарен на уровне значимости 5 %.

Вычисление и оценка показателя Херста с помощью r/s

Для расчета показателя Херста в нестационарном трафике на малых выборках используется *R/S* анализ. Многие исследователи [7–8] применяют *R/S* анализ для нахождения показателя Херста в сетевом трафике. Одно из основных преимуществ *R/S*-анализа заключается в том, что в отличие от многих широко распространенных статистических критериев, он не основан на каких бы то ни было предположениях об организации исходных данных (о том, какому закону распределения они подчиняются). Очень быстрый и легко реализуемый.

Алгоритм *R*/*S*:

$$\frac{R}{S} = (aN)^H, \tag{6}$$

откуда

$$H = \frac{\log(R/S)}{\log(aN)}, \tag{7}$$

где H – показатель Херста; S – среднеквадратичное отклонение ряда наблюдений x; N – число периодов наблюдений; a – заданная константа, положительное число.

$$S = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2}, \tag{8}$$

$\overline{x}$ – среднее арифметическое ряда наблюдений x за N периодов

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{9}$$

Размах накопленного отклонения R это разность между максимальным и минимальным накопленными отклонениями:

$$R = \max_{1 \le u \le N} Z_u - \min_{1 \le u \le N} Z_u, \tag{10}$$

где Z_u – накопленное отклонение ряда x от среднего $\overline{x}$:

$$Z_u = \sum_{i=1}^{u}(x_i - \overline{x}), \tag{11}$$

Из формулы видно, что на рост показателя Херста влияют:

увеличение размаха колебаний *R*;

уменьшение среднеквадратичного отклонения *S*;

уменьшение количества наблюдений *N*.

При $0{,}5 \le H \le 1{,}0$ мы наблюдаем персистентные, или трендоустойчивые ряды. Если ряд возрастает (убывает) в предыдущий период, то вероятно, что он будет сохранять эту тенденцию еще какое-то время в будущем. Наблюдения не являются независимыми. Каждое наблюдение несет память обо всех предшествующих событиях. Процесс обладает длительной памятью. Эта память долговременная, теоретически она сохраняется навсегда. Трендоустойчивость поведения, или сила персистентности, увеличивается при приближении H к единице. Обычно тот факт, что $0{,}5 < H < 1$, считается достаточным основанием для признания процесса самоподобным.

При $H = 0{,}5$ ряд является случайным (последующие значения временного ряда не связаны с его предыдущими значениями).

При $0 < H < 0{,}5$ ряд является антиперсистентным (последующие изменения значений временного ряда противоположны его предыдущему поведению.

Для проверки *R*/*S* сформирован датасет состоящий из легитимного.

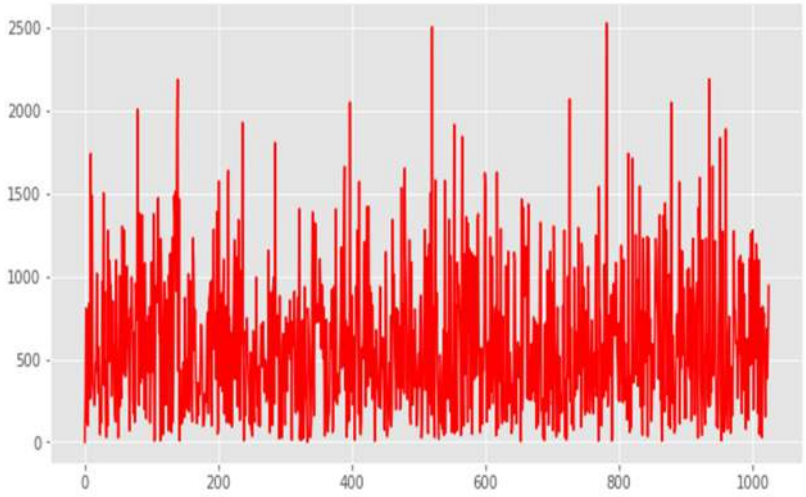

Рис. 2. Легитимный сетевой трафик

После применения *R*/*S* анализа для легитимного трафика, построена логарифмическая регрессия (рис. 3) и вычислен показатель Херста равный 0,56.

Далее *R*/*S* применялся к аномальному сетевому трафику (рис. 4).

Применив *R*/*S* анализ для аномального трафика, построена логарифмическая регрессия (рис. 5) и вычислен показатель Херста равный 1,378.

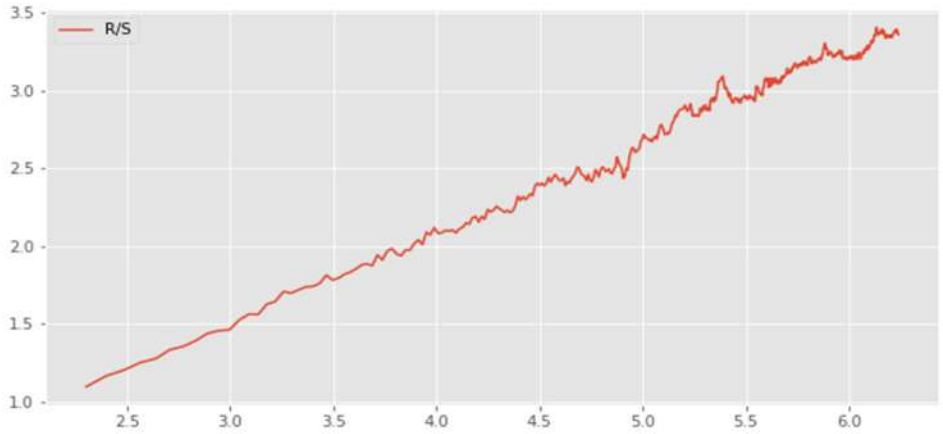

Рис. 3. Зависимость *R*/*S* от времени в логарифмической шкале (Херста = 0,56)

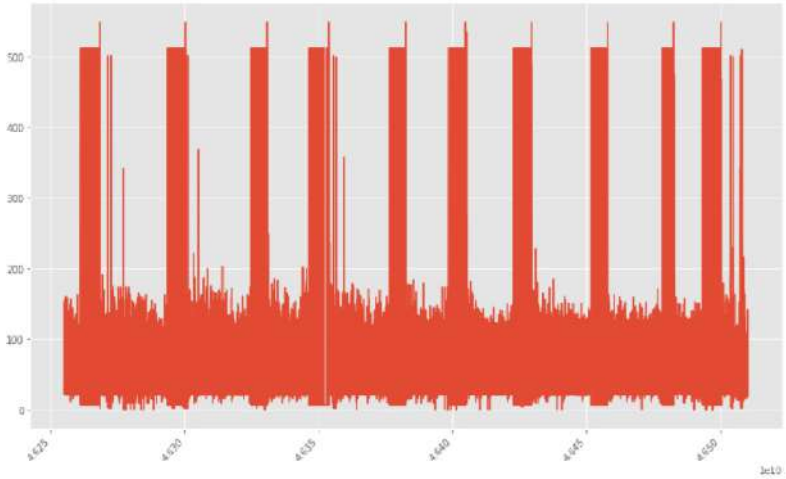

Рис. 4. Аномальный сетевой трафик

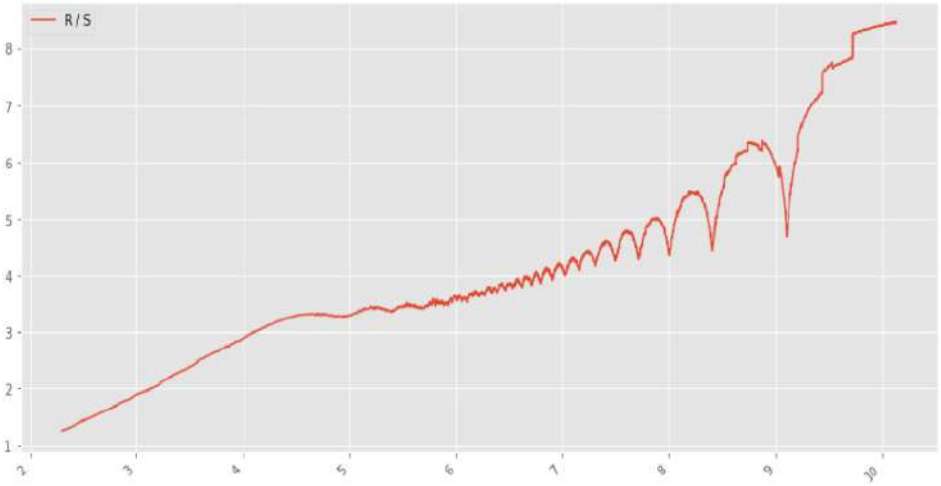

Рис. 5. Зависимость *R*/*S* от времени в логарифмической шкале (Херста = 1,378)

Как видно из рис. 5 показатель Херста превышает максимальное константное значение 1, что подтверждает наличие аномалий в сетевом трафике.

Вычисление и оценка показателя Херста с помощью DFA

Для расчета показателя Херста в нестационарном трафике на зашумленных и больших объемах данных, для более точных вычислений, предпочтительнее использовать DFA анализ:

1. Преобразование временного ряда $x(t)$ в функцию кумулятивных сумм (профиль функции) путем суммирования значений временного ряда:

$$X(t) = \sum_{i=1}^{N} \left(x_i(t) - \overline{x} \right), \tag{12}$$

$$\overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i(t), \tag{13}$$

2. Временной ряд, разбивается на $\frac{N}{n}$ непересекающихся интервалов.

3. В пределах каждого интервала осуществляется линейная аппроксимация ряда $x(t)$ методом наименьших квадратов – выделяется локальный тренд: $y_j(t) = a_j t + b_j$, где a_j и b_j – константы для каждого интервала.

4. Для каждого интервала устраняется локальный тренд путем перехода к разности $X(t) - y_j(t)$ и проводится анализ среднеквадратичного отклонения от локального тренда, т.е. вычисляется функция:

$$F_j^2(n) = \frac{1}{n} \sum_{t=jn+1}^{(j+1)n} \left(X(t) - y_j(t) \right)^2, \tag{14}$$

5. Далее вычисляется среднее значение:

$$F^2(n) = \frac{n}{N} \sum_{j=0}^{\frac{N}{n}-1} F_j^2(n), \tag{15}$$

Если исследуемый ряд сводится к самоподобному множеству, проявляющему дальнодействующие корреляции, то флуктуационная функция $F(n)$ представляется степенной зависимостью:

$$F(n) \sim n^H, \tag{16}$$

где H – показатель Херста.

H может быть вычислен с помощью метода наименьших квадратов как угловой коэффициент прямой, определяющей зависимости $\log F(n)$ от $\log(n)$.

С помощью алгоритма DFA проведен расчет показателя Херста нестационарного трафика, описываемого моделью с заранее заданным показателем Херста = 0,5.

С помощью DFA, построена логарифмическая регрессия и найден показатель Херста, равный 0,49 (рис. 6). Полученный результат почти полностью совпадает с заданной величиной, что подтверждает эффективность текущего метода на зашумленных рядах.

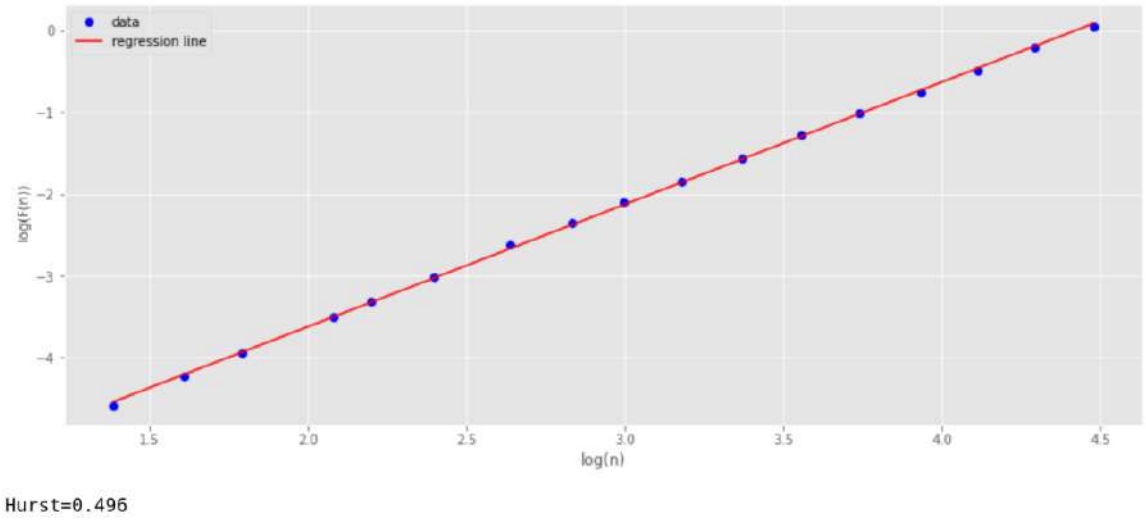

Рис. 6. Зависимость $F(n)$ от времени в логарифмической шкале (Херст = 0,496)

Анализ показал, что метод DFA исключает линейный тренд из каждого анализируемого фрагмента временного ряда, что позволяет повысить точность в условиях низкочастотных помех или на больших объемах данных. В тоже время *R*/*S* является более быстрым алгоритмом вычисления показателя Херста, не уступающим в точности на небольших объемах данных. Поэтому *R*/*S* является предпочтительным для дальнейшего исследования.

Таким образом, эксперименты, проведенные на эталонных выборках, состоящие из легитимного и аномального трафика, продемонстрировали наличие самоподобия трафика КС и возможность достаточно точного определения показателя самоподобия на основе рассмотренных алгоритмов.

Заключение

В статье разработана модель выявления аномалий в сетевом трафике СПД в условиях КА, отличающаяся от известных возможностью описывать стационарный и нестационарный сетевой трафик, классифицировать его и в зависимости от вида трафика обосновать метод по выявлению аномалий.

Модель основана на использовании основных положений теории фракталов и предлагаемых этой теорией методов оценки самоподобия, *R*/*S*-анализ и метод DFA. При тестировании фрактальных методов, позволяющих проводить исследования долговременных зависимостей в трафике КС, метод DFA оказался более эффективен, чем *R*/*S*-анализ на зашумленных данных или больших выборках, из-за исключения линейного тренда из каждого анализируемого фрагмента временного ряда. Следовательно, DFA позволяет обнаруживать корреляции на большие расстояния, встроенные в нестационарные ряды, что характерно для КС, избегая ложного обнаружения явных корреляций на большие расстояния, которые являются артефактами нестационарности. Неоспоримым преимуществом *R*/*S*-анализа являются более быстрые вычисления показателя Херста, а эффективность алгоритма не уступает в точности DFA, на небольших объемах данных.

Основываясь на результатах тестирования, можно сделать вывод, что предложенная модель является достаточно адекватной. Эксперименты показали, что существует характерное время, после которого показатель Херста резко меняется. Это время указывает на объем системной памяти. Экспериментальные результаты также свидетельствуют о том, что самоподобные свойства присущи любому сетевому трафику на канальном уровне модели tcp/ip. При появлении сетевых аномалий, вызванных, например, кибератаками типа DDoS и «сканирование сети и ее уязвимостей», характер этих свойств начинает существенно отличаться от нормального трафика.

Также реализованы программные модули для определения стационарности сетевого трафика с помощью расширенного теста Дики-Фуллера, а также вычисления и оценки показателя Херста с помощью *R*/*S* и DFA.

Список литературы

1. Kotenko I., Saenko I., Lauta O., Kribel A. An approach to detecting cyber attacks against smart power grids based on the analysis of network traffic self-similarity // Energies. 2020. Т. 13, № 19.

2. Kotenko I., Saenko I., Lauta O., Karpov M. Methodology for management of the protection system of smart power supply networks in the context of cyberattacks // Energies. 2021. T. 14, № 18.

3. Ably P., Flandrin P., Taqqu M. S., Veitch D. Self-Similarity and long-range dependence through the wavelet lens // In Theory and Applications of Long Range Dependenc. Boston : Birkhauser Press, 2002. P. 345–379.

4. Canadian Smart Grid Framework. Canadian Electricity Association, March 25, 2010.

5. Крибель А. М., Лаута О. С., Филин А. В., Фень А. С. Метод обнаружения аномалий в сетевом компьютерном трафике на основе нейронной сети с использованием LSTM // Электросвязь. 2021. № 12. С. 43–48.

6. Карпов М. А., Лаута О. С., Коцыняк М. А., Крибель А. М. Подход к управлению системой защиты информационно-телекоммуникационной сети специального назначения // Известия Тульского государственного университета. Технические науки. 2020. № 7. С. 216–226.

7. Federal Office for Information Security, “Protection Profile for the Security Module of a Smart Metering System, V.1.0,” March, 2015.

8. Adnan Anwar, Abdun Mahmood. Cyber Security of Smart Grid Infrastructure // The State of the Art in Intrusion Prevention and Detection Publisher. CRC Press, 2014. doi:10.1201/b16390-9

Для цитирования: Перов Р. А., Скоробогатов С. Ю., Бударин Э. А., Сычужников В. Б. Модель выявления аномалий в сетевом трафике сети передачи данных в условиях компьютерных атак // Безопасность информационных технологий : сб. науч. ст. по материалам IV Всерос. науч.-техн. конф. : в 2 т. Пенза : Изд-во ПГУ, 2022. Т. 1. С. 143–154.

СВЕДЕНИЯ ОБ АВТОРАХ

Аккуратнов Александр Николаевич, аспирант, Пензенский государственный университет, г. Пенза.

Бакмаева Кристина Руслановна, соискатель 32 кафедры, Военная академия связи имени Маршала Советского Союза С. М. Буденного Министерства обороны Российской Федерации, г. Санкт-Петербург.

Банных Андрей Григорьевич, аспирант, Пензенский государственный университет, г. Пенза.

Баринов Даниил Евгеньевич, студент, Пензенский государственный университет, г. Пенза.

Безяев Александр Викторович, к.т.н., ведущий научный сотрудник, Пензенский филиал «НТЦ "Атлас"», г. Пенза.

Боровский Александр Сергеевич, д.т.н., доцент, заведующий кафедрой управления и информатики в технических системах, Оренбургский государственный университет, г. Оренбург.

Бударин Эдуард Альбертович, к.т.н., доцент 32 кафедры, Военная академия связи имени Маршала Советского Союза С. М. Буденного Министерства обороны Российской Федерации, г. Санкт-Петербург.

Волчихин Владимир Иванович, д.т.н., профессор, президент Пензенского государственного университета, заслуженный деятель науки РФ, г. Пенза.

Вятчанин Сергей Евгеньевич, к.т.н., доцент, начальник кафедры Военного учебного центра, Пензенский государственный университет, г. Пенза.

Евтихин Иван Олегович, соискатель 32 кафедры, Военная академия связи имени Маршала Советского Союза С. М. Буденного Министерства обороны Российской Федерации, г. Санкт-Петербург.

Зефиров Сергей Львович, к.т.н., доцент, заведующий кафедрой информационной безопасности систем и технологий, Пензенский государственный университет, г. Пенза.

Золотарева Татьяна Александровна, старший преподаватель кафедры информатики, информационных технологий и защиты информации, Липецкий государственный педагогический университет имени П. П. Семенова-Тян-Шанского, г. Липецк.

Иванов Александр Иванович, д.т.н., профессор, научный консультант, Пензенский научно-исследовательский электротехнический институт, г. Пенза.

Иванов Алексей Петрович, к.т.н., доцент, заведующий кафедрой технических средств информационной безопасности, Пензенский государственный университет, г. Пенза.

Качалин Сергей Викторович, к.т.н., заместитель начальника отдела, «НПП "Рубин"», г. Пенза.

Кильдюшкин Кирилл Олегович, студент, Пензенский государственный университет, г. Пенза.

Князьков Владимир Сергеевич, д.т.н., профессор, главный научный сотрудник, Научно-исследовательский институт фундаментальных и прикладных исследований, Пензенский государственный университет, г. Пенза.

Куприянов Евгений Николаевич, аспирант, Пензенский государственный университет, г. Пенза.

Лаута Олег Сергеевич, д.т.н., ст. преподаватель 32 кафедры, Военная академия связи имени Маршала Советского Союза С. М. Буденного Министерства обороны Российской Федерации, г. Санкт-Петербург.

Лукин Виталий Сергеевич, ассистент кафедры информационной безопасности систем и технологий, Пензенский государственный университет, г. Пенза.

Малыгин Александр Юрьевич, д.т.н., профессор кафедры радио- и спутниковой связи Военного учебного центра, Пензенский государственный университет, г. Пенза.

Малыгина Елена Александровна, к.т.н., докторант кафедры технических средств информационной безопасности, Пензенский государственный университет, г. Пенза.

Олейник Юрий Иванович, к.т.н., главный специалист, Радиозавод, г. Пенза.

Панфилова Ирина Евгеньевна, преподаватель кафедры электронных систем и информационной безопасности, Самарский государственный технический университет, г. Самара.

Перфилов Константин Александрович, преподаватель-исследователь межотраслевой лаборатории тестирования биометрических устройств и технологий, Пензенский государственный университет, г. Пенза.

Перов Роман Александрович, адъюнкт 32 кафедры, Военная академия связи имени Маршала Советского Союза С. М. Буденного Министерства обороны Российской Федерации, г. Санкт-Петербург.

Полковникова Светлана Андреевна, аспирант, Пензенский государственный университет, г. Пенза.

Ракицкий Станислав Николаевич, к.т.н., профессор 32 кафедры, Военная академия связи имени Маршала Советского Союза С. М. Буденного Министерства обороны Российской Федерации, г. Санкт-Петербург.

Савинов Константин Николаевич, старший преподаватель кафедры проводной электросвязи и автоматизированных систем, Военный учебный центр, Пензенский государственный университет, г. Пенза.

Серикова Юлия Игоревна, аспирант, Пензенский государственный университет, г. Пенза.

Скоробогатов Сергей Юрьевич, адъюнкт 32 кафедры, Военная академия связи имени Маршала Советского Союза С. М. Буденного Министерства обороны Российской Федерации, г. Санкт-Петербург.

Спирин Сергей Владимирович, к.т.н., старший преподаватель 2 кафедры, Военная академия связи имени Маршала Советского Союза С. М. Буденного Министерства обороны Российской Федерации, г. Санкт-Петербург.

Сулавко Алексей Евгеньевич, к.т.н., доцент кафедры комплексной защиты информации, Омский государственный технический университет, г. Омск.

Сумин Владислав Алексеевич, студент, Пензенский государственный университет, г. Пенза.

Сычужников Виктор Борисович, к.т.н., старший научный сотрудник Научно-исследовательского центра, Военная академия связи имени Маршала Советского Союза С. М. Буденного Министерства обороны Российской Федерации, г. Санкт-Петербург.

Туреев Сергей Васильевич, к.т.н., директор, Концерн «Созвездие», г. Воронеж.

Цимбал Владимир Анатольевич, заслуженный деятель науки РФ, д.т.н., профессор кафедры автоматизированных систем управления, Филиал Военной академии Ракетных войск стратегического назначения имени Петра Великого, г. Серпухов Московской области.

СОДЕРЖАНИЕ

Научное издание

БЕЗОПАСНОСТЬ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

Сборник научных статей по материалам
IV Всероссийской научно-технической конференции
(г. Пенза, 3 июля 2022 г.)

В двух томах

Том 1

Статьи печатаются в авторской редакции.

Корректор *В. В. Чувашова*
Компьютерная верстка *М. Б. Жучковой*
Дизайн обложки *И. В. Шваревой*

Подписано в печать 26.12.2022.
Формат $60\times84^{1}/_{16}$. Усл. печ. л. 9,42.
Заказ № 831. Тираж 30.

Издательство ПГУ
440026, г. Пенза, ул. Красная, 40
Тел./факс: (8412) 666-049, 666-777; e-mail: iic@pnzgu.ru