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» (ИСВЧПЭ РАН), 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;
ООО «НПФ "КРУГ"», г. Пенза;
ООО «Научно-производственное предприятие "БиоКрипт"», г. Пенза;
ООО «АЛГОМАТ», г. Калининград

Безопасность информационных технологий

Сборник научных статей по материалам
V Всероссийской научно-технической конференции,
посвященной 70-летнему юбилею АО «НПП "Рубин"»

(г. Пенза, 27 сентября 2023 г.)

В двух томах

Том 1

Пенза • Издательство ПГУ • 2023

УДК 681.322
Б39

Безопасность информационных технологий : сб. науч. ст. по материалам V Всерос. науч.-техн. конф., посвящ. 70-летнему юбилею АО «НПП "Рубин"» (г. Пенза, 27 сентября 2023 г.) : в 2 т. – Пенза : Изд-во ПГУ, 2023. – Т. 1. – 218 с.

Б39

ISBN 978-5-907807-02-0

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

Приказ
о подготовке и проведении Всероссийской научно-технической конференции «Безопасность информационных технологий»
№ 724/о от 29.06.2023

ISBN 978-5-907807-02-0

Состав оргкомитета научно-технической конференции:

Председатель – В. И. Волчихин, д.т.н., профессор,
президент ФГБОУ ВО «Пензенский государственный университет».
Сопредседатель – А. А. Тарасов, к.т.н., ген. директор АО «НПП "Рубин"».

Члены оргкомитета: **О. С. Авсентьев**, д.т.н., профессор, профессор кафедры информационной безопасности Воронежского института МВД России (г. Воронеж); **А. В. Безяев** – к.т.н., ведущий научный сотрудник Пензенского филиала АО «НТЦ "Атлас"» (г. Пенза); **А. С. Боровский** – д.т.н., доцент, зав. кафедрой «Управление и информатика в технических системах» ФГБОУ ВО «Оренбургский государственный университет» (г. Оренбург); **М. М. Бутаев** – д.т.н., профессор, ученый секретарь НТС АО «НПП "Рубин"» (г. Пенза); **С. А. Гамкрелидзе** – д.т.н., профессор, директор ФГАНУ «Институт сверхвысокочастотной полупроводниковой электроники имени В. Г. Мокерова Российской академии наук» (ИСВЧПЭ РАН) (г. Москва); **С. Л. Зефиров** – к.т.н., доцент, зав. кафедрой «Информационная безопасность систем и технологий» ФГБОУ ВО «Пензенский государственный университет» (г. Пенза); **В. Ю. Егоров** – к.т.н., начальник отделения ООО «НТП "Криптософт"» (г. Пенза); **Н. А. Егорова** – д.т.н., доцент кафедры «Информационная безопасность систем и технологий» ФГБОУ ВО «Пензенский государственный университет» (г. Пенза); **А. И. Иванов** – д.т.н., профессор, научный консультант АО «ПНИЭИ» (г. Пенза); **А. П. Иванов** – к.т.н., доцент, зав. кафедрой «Технические средства информационной безопасности» на базе АО «ПНИЭИ» (г. Пенза); **В. А. Иванов** – д.т.н., профессор, ген. директор ООО «АЛГОМАТ» (г. Калининград); **С. В. Качалин** – к.т.н., зам. начальника отделения АО «НПП "Рубин"» (г. Пенза); **Г. В. Козлов** – д.т.н., профессор, директор ПИ ФГБОУ ВО «Пензенский государственный университет» (г. Пенза); **Б. В. Костров** – д.т.н., профессор, зав. кафедрой «Электронные вычислительные машины» ФГБОУ ВО «Рязанский радиотехнический университет» (г. Рязань); **И. Д. Королев** – д.т.н., профессор, ФГКВОУ ВО «Краснодарское высшее военное училище имени генерала армии С. М. Штеменко» Министерства обороны Российской Федерации (г. Краснодар); **В. С. Князьков** – д.т.н., профессор, директор Центра научно-исследовательского института фундаментальных и прикладных исследований ФГБОУ ВО «Пензенский государственный университет» (г. Пенза); **В. П. Кулагин** – д.т.н., профессор, зав. кафедрой «Аппаратное, программное и математическое обеспечение вычислительных систем» Института комплексной безопасности и специального приборостроения (г. Москва); **В. М. Лазарев –** д.т.н., профессор, руководитель управления координации научно-технического развития АО «Системы управления» (г. Москва); **О. С. Лаута** – д.т.н., начальник НИЦ ФГКВОУ ВО «Военная академия связи имени Маршала Советского Союза С. М. Буденного» Министерства обороны Российской Федерации (г. С.-Петербург); **А. Ю. Малыгин** – д.т.н., профессор, профессор кафедры «Радио- и спутниковой связи» Военного учебного центра при ФГБОУ ВО «Пензенский государственный университет» (г. Пенза); **Ю. И. Мамон** – 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); **А. А. Привалов –** д.в.н., профессор, академик РАЕН, профессор кафедры «Электрическая связь» ФГБОУ ВО «Петербургский государственный университет путей сообщения императора Александра I» (г. С.-Петербург); **В. А. Пушкин** – к.т.н., доцент, руководитель проекта проектного офиса АО «Радиозавод» (г. Пенза); **А. П. Ремонтов** – к.т.н., доцент, декан факультета информационных и образовательных технологий ФГБОУ ВО «Пензенский государственный технологический университет» (г. Пенза); **В. А. Тихомиров** – д.т.н., профессор, заслуженный деятель науки РФ, генеральный директор АО НПО «Развитие Инновационных Технологий» (г. Тверь); **Д. А. Урядов** – зам. главного конструктора ФГУП ФНПЦ «ПО "Старт" имени М. В. Проценко» (г. Заречный, Пензенская обл.); **В. А. Фунтиков** – к.т.н., генеральный директор АО «ПНИЭИ» (г. Пенза); **П. Н. Цибизов** – к.т.н., доцент, начальник НИО АО «НИЭФИ» (г. Пенза); **В. А. Цимбал –** д.т.н., профессор, заслуженный деятель науки РФ, филиал ВА имени Петра Великого (г. Серпухов); **Ю. К. Язов** – д.т.н., профессор, главный научный сотрудник Управления ФАУ «ГНИИИ ПТЗИ ФСТЭК России» (г. Воронеж)

УДК 9:004.38:621.3:623

ПУТЬ ОТ ПЕРВОЙ СЕРИЙНОЙ ЭВМ ДО КРУПНЫХ ТЕРРИТОРИАЛЬНО РАСПРЕДЕЛЕННЫХ СЛОЖНЫХ АВТОМАТИЗИРОВАННЫХ СИСТЕМ И КОМПЛЕКСОВ УПРАВЛЕНИЯ СПЕЦИАЛЬНОГО НАЗНАЧЕНИЯ

А. А. Тарасов

Научно-производственное предприятие «Рубин», г. Пенза

Аннотация. Представлена история предприятия под научно-техническим руководством Б. И. Рамеева. Отмечаются достижения динамично развивающегося предприятия, специализирующегося на создании сложных территориально распределенных автоматизированных систем управления и комплексов управления специального назначения. Рассмотрены научные школы предприятия, перспективы предприятия.

Ключевые слова: электронные вычислительные машины, автоматизированные системы управления, Научно-производственное предприятие «Рубин»

THE WAY FROM THE FIRST SERIAL COMPUTER TO LARGE GEOGRAPHICALLY DISTRIBUTED COMPLEX AUTOMATED SYSTEMS AND SPECIAL-PURPOSE CONTROL COMPLEXES

A. A. Tarasov

Scientific-industrial Enterprise «Rubin», Penza

Abstract. History of the enterprise under the scientific and technical direction of B. I. Rameev. Achievements of a dynamically developing enterprise specializing in the creation of complex geographically distributed automated control systems and special-purpose control complexes. Scientific schools of the enterprise. Prospects of the enterprise.

Keywords: electronic computers, automated control systems, Scientific-industrial Enterprise «Rubin»

В силу исторически сложившихся обстоятельств акционерное общество «Научно-производственное предприятие «Рубин» стояло у истоков процесса создания и развития аналоговой и цифровой вычислительной техники в СССР. Свидетельством тому является первая в СССР серийная ЭВМ «Урал-1», разработанная под руководством талантливого ученого и конструктора, лауреата

Сталинской премии и премии Совета Министров СССР, заслуженного изобретателя СССР Башира Искандаровича Рамеева.

Б. И. Рамеев по праву занимает почетное место в ряду ученых – основоположников компьютерной науки, а его имя, известное во всем мире, тесно связано с историей АО «НПП «Рубин» [1–3].

В 1955 году, переехав в Пензу с группой талантливых молодых специалистов, Б. И. Рамеев стал главным инженером и заместителем директора по научной работе п/я 24 (Научно-исследовательский институт математических машин, а ныне – АО «НПП «Рубин»). Коллектив Б. И. Рамеева разработал и запустил в производство 15 типов универсальных и специализированных вычислительных машин, а также более 100 различных периферийных устройств.

Уже первая ЭВМ «Урал», выпущенная в Пензе в 1957 году, использовалась во многих вычислительных центрах СССР (так же, как и созданная в те годы ламповая машина IBM 650 широко эксплуатировалась в американской промышленности). Именно эта ЭВМ стала экспонатом Политехнического музея в Москве. Газета Computer World (Россия) написала об этом событии в номере от 28 октября 1999 года: «*Эта машина не зря удостоилась чести быть выставленной в музее наших достижений... Это была недорогая и вполне эффективная для инженерных расчетов машина, благодаря чему и пользовалась большим спросом... Первый «Урал», и его последователи долгие годы верой и правдой служили советским инженерам...*» [4].

Во время пензенского этапа (1955–1968 гг.) научно-технической деятельности под руководством выдающегося ученого были разработаны и приняты в промышленное производство ЭВМ «Стрела», серия ЭВМ «Урал», специализированные ЭВМ «Погода», «Гранит», «Кристалл» и др. В 1960–1970-х годах транзисторные «Уралы» («Урал-11», «Урал-14», «Урал-16»), а их было выпущено более 500, работали в каждом втором вычислительном центре страны, в ЦУП на Байконуре, ОАСУ Госбанк, «Строитель», «Листопрокат» [5].

Б. И. Рамеев внес огромный вклад в становление и развитие тематической направленности предприятия, создание пензенской научной школы конструирования средств вычислительной техники и систем управления.

В 2013 году, в год 60-летия предприятия, в АО «НПП «Рубин» учреждена премия им. Б. И. Рамеева за вклад в развитие приоритетных направлений науки и техники, крупные научные разработки,

имеющие большое значение для науки и практики, особые заслуги в создании и внедрении принципиально новой высокоэффективной техники и технологии.

К настоящему времени коллективом АО «НПП «Рубин» пройден славный путь от создания ламповых и полупроводниковых ЭВМ, первых информационно-управляющих систем и комплексов, использующихся в энергетике, металлургии, строительстве, до автоматизированных систем и комплексов управления специального назначения.

Сложность задачи и комплексный подход к работе позволили сформировать коллектив высококвалифицированных специалистов, включающий инженеров, программистов, конструкторов, технологов и рабочих высокого разряда. Это дало возможность предприятию одновременно с разработкой устройств и машин вести работы по созданию управляющих и вычислительных комплексов с применением ЭВМ в интересах народного хозяйства и Министерства обороны СССР. В дальнейшем предприятие выполняло работы сразу по нескольким направлениям радиоэлектронного профиля и таким образом поддерживало, и укрепляло звание разработчика радиоэлектронных устройств на протяжении всей своей истории.

За 70-летний период научно-производственной деятельности АО «НПП «Рубин» прошло испытание «перестройкой», трудности 1990-х годов и неопределенности начала XXI века. Предприятие трансформировалось, когда это было необходимо, но двигалось вперед и развивалось.

В 1970-х и 1980-х годах создана и введена в эксплуатацию отраслевая АСУ Госбанка СССР, включавшая в себя более 90 вычислительных центров в крупных городах страны, вычислительный комплекс «Галета» системы фотограмметрической обработки космических снимков «Аналит». Эти работы принесли институту всесоюзное признание, а коллективу – высокие государственные награды. За достигнутые успехи в создании новых сложных образцов систем и комплексов предприятие было награждено орденом Трудового Красного Знамени, 16 специалистов удостоены звания лауреата Государственной премии СССР, премии Госкомоборонпрома РФ в области науки и техники, более 700 сотрудников награждены орденами и медалями СССР и Российской Федерации.

В турбулентные 1990-е годы предприятие смогло сохранить и расширить тематику работ, найти эффективные варианты

использования «двойных» технологий, что способствовало усилению позиций в создании стационарно-мобильных систем и комплексов для министерств и ведомств Российской Федерации, а в ряде направлений стать одним из ведущих предприятий страны.

В 2004 г. Указом Президента Российской Федерации АО «НПП «Рубин» было включено в состав ОАО «Концерна радиостроения «Вега». В течение последующих лет предприятие формировалось как научно-производственное: на предприятии были образованы три научно-технических центра по направлениям деятельности, конструкторское отделение, базовый центр системного проектирования, создан производственный комплекс, оснащенный самым современным производственно-технологическим и инструментальным оборудованием.

В 2000–2010-х годах коллективом предприятия разработаны новейшие стационарные и мобильные комплексы управления и связи, разведки, тылового и технического обеспечения, организованы их серийное производство и поставка в войска и органы военного управления.

Крупнейшими опытно-конструкторскими работами стали ОКР «Цезарь-2» и ОКР «Премьер». В рамках ОКР «Цезарь-2» разработаны аппаратно-программные стационарные и мобильные комплексы оперативных штабов, которые были поставлены во все регионы страны. В рамках ОКР «Премьер» созданы технологическая система динамического воспроизведения и обработки полетной информации авиационного комплекса дальнего радиолокационного обнаружения, технологический аппаратно-программный комплекс регистрации и синтеза радиотехнических сигналов бортовой обзорной РЛС авиационного комплекса дальнего радиолокационного обнаружения, а также автоматизированное рабочее место инженера-оператора РЛС для решения задач анализа состояния и управления РЛС авиационного комплекса дальнего радиолокационного обнаружения А-50У.

Еще одной крупной работой, заслужившей высокую оценку заказчика, стало создание базового комплекта подсистемы управления ПВО единой системы управления войсками (оружием) в тактическом звене «Барнаул-Т», ориентированного на автоматизацию процесса управления силами и средствами ПВО тактического звена в единой системе от стрелка-зенитчика до начальника ПВО соединений в ходе подготовки и ведения противовоздушного боя. С разработкой базового комплекта «Барнаул-Т» в Вооруженных Силах Российской Федерации у предприятия появилась

экспортно привлекательная продукция и началась новая страница истории АО «НПП «Рубин».

АО «НПП «Рубин» не раз было отмечено государственными премиями. Два года подряд, в 2013 и 2014 годах, коллективы разработчиков получали Государственную премию Российской Федерации имени Маршала Советского Союза Г. К. Жукова. За большой вклад в области создания вооружения и военной техники в 2013 году авторский коллектив был награжден Государственной премией Российской Федерации имени Маршала Советского Союза Г. К. Жукова. Руководство страны высоко оценило комплексную работу по созданию на базе программно-технического комплекса «Барнаул-Т» автоматизированной интегрированной гетерогенной системы противовоздушной обороны тактического звена, способствующую значительному укреплению обороноспособности Российской Федерации. В 2014 году сотрудники АО «НПП «Рубин» вновь были удостоены высокой государственной награды. Указом Президента Российской Федерации В. В. Путина была присуждена Государственная премия Российской Федерации имени Маршала Советского Союза Г. К. Жукова в области создания вооружения и военной техники за аппаратно-программные комплексы оперативных штабов в субъектах Российской Федерации по управлению выделенными силами и средствами для проведения контртеррористической операции, способствующие эффективному решению проблем национальной безопасности государства.

Сотрудники АО «НПП «Рубин» были отмечены Национальной премией «Золотая идея» Федеральной службы военно-технического сотрудничества за вклад в экспортную продукцию. Показателем высокой оценки специалистов АО «НПП «Рубин» стало неоднократное присвоение коллективам разработчиков стипендии Минпромторга России за выдающиеся достижения в создании прорывных технологий и разработку современных образцов вооружения, военной и специальной техники в интересах обеспечения обороны страны и безопасности государства.

Сегодня АО «НПП «Рубин» – современное, динамично развивающееся предприятие, российской радиоэлектронной промышленности, специализирующееся на создании сложных и производство территориально распределенных автоматизированных систем и комплексов управления специального назначения, информационных систем. АО «НПП «Рубин» выполняет работы по всему жизненному циклу создаваемой продукции: НИР, ОКР,

серийный выпуск изделий, авторский и технический надзор в процессе эксплуатации, ремонт и утилизация изделий. Как головной исполнитель ряда государственных контрактов предприятие ведет разработку и серийную поставку продукции для Министерства обороны, Министерства внутренних дел и других силовых ведомств Российской Федерации. Предприятие выполняет работы на основе концепции комплексного подхода к решению проблем автоматизации управления и является системным интегратором в области создания мобильных комплексов управления и телекоммуникационных средств, работая в тесном взаимодействии с ведущими предприятиями концернов «Вега», «Созвездие», «Автоматика» и др.

За всем этим стоят талантливые и преданные люди, которые являются частью одной профессиональной команды.

АО «НПП «Рубин» – высокотехнологичное предприятие с большим научно-техническим потенциалом, научными школами по ряду перспективных направлений:

– системное проектирование крупных территориально распределенных информационно-управляющих систем специального назначения стационарного и мобильного базирования;

– создание интегрированных автоматизированных систем управления материально-техническим обеспечением Вооруженных Сил Российской Федерации;

– создание систем автоматизированного управления противовоздушной обороной тактического звена;

– разработка распределенных телекоммуникационных систем и мобильных пунктов управления специального назначения.

Предприятие занимает достойное место в радиоэлектронной отрасли, тем не менее никогда не останавливается на достигнутом. Реализовать этот принцип позволяет планомерное выполнение программы научно-технического развития акционерного общества, а также участие в федеральных целевых программах.

Участие в ФЦП «Развитие электронной компонентной базы и радиоэлектроники», «Развитие оборонно-промышленного комплекса Российской Федерации» позволило реконструировать производственные площади, модернизировать технологические процессы, заменить устаревшее оборудование, улучшить условия труда. В результате предприятие имеет практически полный комплекс современного оборудования для выпуска технически сложной продукции как военного, так и гражданского назначения.

Изделия АО «НПП «Рубин» создаются с учетом последних достижений отечественной науки в области оптоэлектроники, радиолокации, радиотехники, разведки и связи. Разработка составных частей стационарных и мобильных автоматизированных систем и комплексов управления, программного и информационного обеспечения ведется с использованием постоянно развиваемых аппаратно-моделирующих комплексов главных конструкторов, включающих полунатурные и натурные модели создаваемых изделий и аппаратно-программные технологические средства разработки программного и информационно-лингвистического обеспечения.

В АО «НПП «Рубин» созданы все предпосылки для того, чтобы продукция и услуги предприятия были востребованы в течение нескольких последующих лет на рынке, а также выработан план действий по обновлению своих технологий, расширению номенклатуры выпускаемой продукции, круга заказчиков и межведомственной кооперации, наращиванию экспортных возможностей. Предприятие участвует в реализации Указа Президента Российской Федерации «О развитии искусственного интеллекта в РФ» № 480 от 19.10.19, входит в состав членов технического комитета № 164 «Искусственный интеллект», в подкомитет, занимающийся проблемами разработки доверенного искусственного интеллекта. Предприятие активно включилось в программы диверсификации промышленности и развивает несколько направлений по выпуску высокотехнологичной гражданской продукции и продукции двойного назначения.

Все прилагаемые усилия позволяют АО «НПП «Рубин» соблюдать в работе баланс между специальными стратегическими целями по эффективному выполнению ответственных государственных заданий, научно-техническому развитию и коммерческими целями по созданию высокой конкурентоспособности на рынках высокотехнологичной продукции.

Предприятие, базирующееся на мощной научной школе, никогда не отделяет теоретические исследования от практики, что обеспечивает мировой уровень его продукции. Технический базис, исторические традиции научного поиска, благодаря которым существует и развивается АО «НПП «Рубин» в наши дни, позволяют предприятию непрерывно расти и уверенно планировать свое развитие. АО «НПП «Рубин», как и прежде, устремлено в будущее!

Список литературы

1. Страницы истории отечественных ИТ / сост. Э. М. Пройдаков. М. : Альпина Паблишер, 2015. Т. 1. 265 с.

2. Малиновский Б. Н. История вычислительной техники в лицах. Киев, 1995. 384 с.

3. История отечественной радиоэлектронной техники для сухопутных войск. М. : Столичная энциклопедия, 2018. 520 с.

4. Ware W. H. Soviet Computer Technology-1959 // IRE Transactions on electronic computers. 1960. March. P. 72–120.

5. Голубинцев В. О. [и др.]. Эволюция универсальных вычислительных машин. М. : Сов. радио, 1980.

УДК 004.056

ОБЕСПЕЧЕНИЕ БЕЗОПАСНОСТИ ПОЛЬЗОВАТЕЛЬСКИХ ДАННЫХ СТАНДАРТНЫМИ СРЕДСТВАМИ ЯЗЫКА ПРОГРАММИРОВАНИЯ PYTHON НА ПРИМЕРЕ ПРОЦЕССА РЕГИСТРАЦИИ И АВТОРИЗАЦИИ ПОЛЬЗОВАТЕЛЕЙ ВЕБ-САЙТОВ

А. В. Архипов[1], К. М. Демушкина[2], М. О. Демушкин[3], А. В. Казбаев[4]

[1,2,3,4] *Научно-производственное предприятие «Рубин», г. Пенза*

Аннотация. На сегодняшний день кибератаки являются одними из наиболее популярных методов мошенничества. Наиболее подверженной средой к взлому пользовательских данных являются веб-сайты госучреждений. На 2022 г. доля атак на веб-сайты госучреждений составила 41 % от общего числа атак. Очевидно, что задача обеспечения целостности пользовательских данных является актуальной на сегодняшний день. Существуют методы и способы обеспечения безопасной работы на веб-сайтах, однако все они, так или иначе, требуют дополнительного оборудования или специальных библиотек и модулей при разработке. Поэтому основная цель заключается в исследовании возможности обеспечения безопасности пользовательских данных стандартными средствами языка программирования Python на примере процесса аутентификации и регистрации пользователей веб-сайтов. В контексте обеспечения безопасности пользовательских данных средствами языка программирования Python разработчики рассматривают библиотеки RSA и PyCrytodome. Однако RSA имеет ограниченный алгоритм шифрования, а PyCrytodome требует скачки и установки в проект. Для обеспечения целостности данных можно использовать стандартные библиотеки Python: hmac, hashlib, base64. Данные библиотеки обеспечат необходимый уровень целостности пользовательских данных, а разработанные алгоритмы являются эффективным способом обеспечения безопасности хранения и передачи данных при аутентификации и регистрации пользователей на веб-сайте.

Ключевые слова: безопасность данных, веб-сайт, python, хеширование, регистрация, аутентификация

ENSURING THE SECURITY OF USER DATA BY STANDARD MEANS OF THE PYTHON PROGRAMMING LANGUAGE ON THE EXAMPLE OF THE PROCESS OF REGISTRATION AND AUTHORIZATION OF WEBSITE USERS

A. V. Arkhipov[1], K. M. Demushkina[2], M. O. Demushkin[3], A. V. Kazbaev[4]

[1,2,3,4] *Scientific-industrial Enterprise «Rubin», Penza*

Abstract. Today cyberattacks are one of the most popular methods of fraud. The most vulnerable environment for hacking user data are the websites of government

agencies. By 2022, the share of attacks on the websites of state institutions amounted to 41 % of the total number of attacks. Obviously, the task of ensuring the integrity of user data is relevant today. There are methods and ways to ensure safe operation on websites, but all of them, one way or another, require additional equipment or special libraries and modules during development. Therefore, the main goal is to investigate the possibility of ensuring the security of user data by standard means of the Python programming language using the example of the authentication and registration process of website users. In the context of ensuring the security of user data by means of the Python programming language, developers consider the RSA and PyCrytodome libraries. However, RSA has a limited encryption algorithm, and PyCrytodome requires downloading and installing into the project. To ensure data integrity, you can use standard Python libraries: hmac, hashlib, base64. These libraries will provide the necessary level of user data integrity, and the developed algorithms are an effective way to ensure the security of data storage and transmission during authentication and registration of users on the website.

Keywords: data security, website, python, hashing, registration, authentication

Кибератаки остаются одним из наиболее распространенных видов мошенничества в мире. Только за 2022 год количество успешных атак возросло на 20,8 % по сравнению с предыдущим годом. В связи с массовыми утечками данных появляется возможность проведения атак с использованием скомпрометированной информации о пользователях. Число успешных атак, направленных на веб-ресурсы организаций, увеличилось на 56 %. Если в 2021 году веб-ресурсы компаний становились объектами атак в 17 % случаев, то в 2022 году доля таких инцидентов составила 22 %. С ростом количества кибератак столкнулись организации многих отраслей, наибольший удар пришелся на госучреждения: количество инцидентов выросло более чем в 2 раза, а их доля повысилась с 23 % до 41 %. Авторы статьи «Актуальные киберугрозы: итоги 2022 года» отмечают, что: «Увеличение числа атак на веб-ресурсы также обусловлено появлением уязвимостей, найденных, в частности, в популярных плагинах, таких как WordPress, Magento (плагин для e-commerce)» [1].

Поэтому, в настоящее время проблема безопасности пользовательских данных остается актуальной задачей. Существует множество способов для решения проблемы утечки данных пользователей. Чаще всего способы зависят от языка программирования, на котором разрабатывается программное обеспечение [2, 3]. В данной статье будет рассматриваться задача обеспечения безопасности пользовательский данных веб-сайтов. стандартными библиотеками языка Python.

Один из путей решений проблем безопасности данных это использование библиотек RSA и PyCrytodome [4]. Библиотека RSA – это библиотека для шифрования и дешифрования данных с использованием алгоритма RSA. RSA (Rivest-Shamir-Adleman) – это криптографический алгоритм, который использует пару ключей: открытый ключ для шифрования данных и закрытый ключ для их дешифрования. RSA является одним из самых популярных алгоритмов шифрования, используемых в современных системах безопасности. PyCryptodome также предоставляет функции для генерации хешей данных. Хеширование – это процесс преобразования данных фиксированной длины, который позволяет быстро проверить целостность данных. PyCryptodome поддерживает алгоритмы хеширования, такие как SHA-256 (Secure Hash Algorithm 256-bit) и MD5 (Message Digest Algorithm 5) [5]. В основном за основу обеспечения безопасности используются библиотеки, рассмотренные выше, однако иногда разработчикам приходится работать только стандартными библиотеками Python без возможности установки и настройки дополнительных библиотек. Отсюда возникает задача: обеспечение целостности пользовательских данных стандартными средствами языка программирования Python.

За основу инструментов для разработки можно взять модуль hmac и библиотеку hashlib для работы с хешированием и base64 для кодировки. Модуль hmac представляет собой алгоритм HMAC, который использует хеш-функции. Данный модуль позволяет проверить целостность информации, а также гарантирует, что данные, передаваемые или хранящиеся в ненадежной среде, не были изменены. Алгоритмы HMAC описывают стандарты обмена данными с проверкой целостности по секретному ключу.

Библиотека hashlib представляет собой общие методы для безопасных алгоритмов хеширования. В ее основу входит множество алгоритмов хеширования: FIPS SHA1, SHA224, SHA256, SHA384 и SHA512, определенные в FIPS 180-2, а также алгоритм MDA RSA, определенный в Интернете RFC 1321.

Библиотека base64 позволяет осуществлять кодирование и декодирование данных. Этот модуль предоставляет два интерфейса. Современный интерфейс поддерживает кодирование байтовоподобных объектов в байты ASCII и декодирование байтообразных объектов или строк, содержащих ASCII в байты. Поддерживаются оба алфавита base-64, определенные в RFC 3548 –

это обычный и безопасный для URL и файловой системы. Рассмотрим обеспечение безопасности данных в момент регистрации пользователя на сайте, алгоритм взаимодействия модулей представлен на рис. 1.

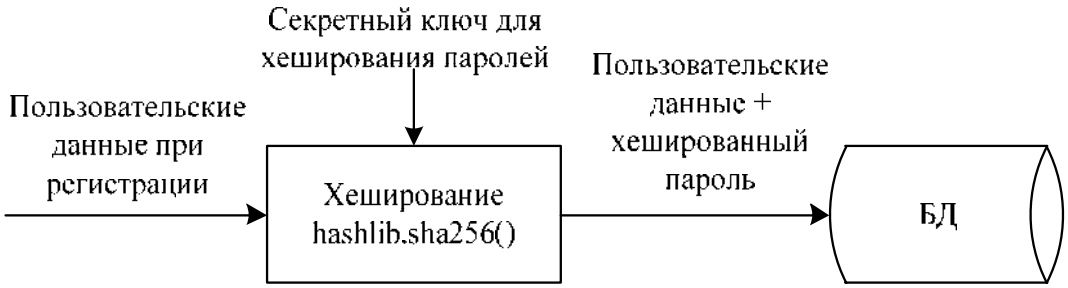


Рис. 1. Алгоритм хеширования данных
при регистрации пользователей

В момент регистрации пользователь передает свои личные данные серверу (логин/пароль/ФИО). Задача системы сделать хранение данных в безопасном виде. Для этого будет использоваться библиотека hashlib. Данные от пользователя поступают на сервер, перед регистрацией данных в базе, пароль хешируется методом hashlib.sha256(). Функция hashlib.sha256 использует алгоритм SHA-256, который является одним из наиболее распространенных алгоритмов хеширования. SHA-256 генерирует хеш-значение длиной 256 бит (32 байта) и обеспечивает высокий уровень безопасности. Важно отметить, что хеширование паролей осуществляется с использованием секретного ключа, что повышает целостность пользовательских данных. Секретный ключ хранится у администратора сайта и доступен лишь правообладателю, поэтому такой алгоритм хеширования повышает безопасность данных. Данные пользователя регистрируются в базе, с учетом того, что пользовательские пароли хранятся в хешированном виде.

Таким образом, это дает возможность разработчикам избежать проблемы «слития» базы данных со всеми паролями и логинами пользователей веб-сайтов, поскольку утечка данных не позволит злоумышленникам воспользоваться полученными знаниями.

Алгоритм обеспечения безопасности пользовательских данных в момент авторизации на сайте представлен на рис. 2.

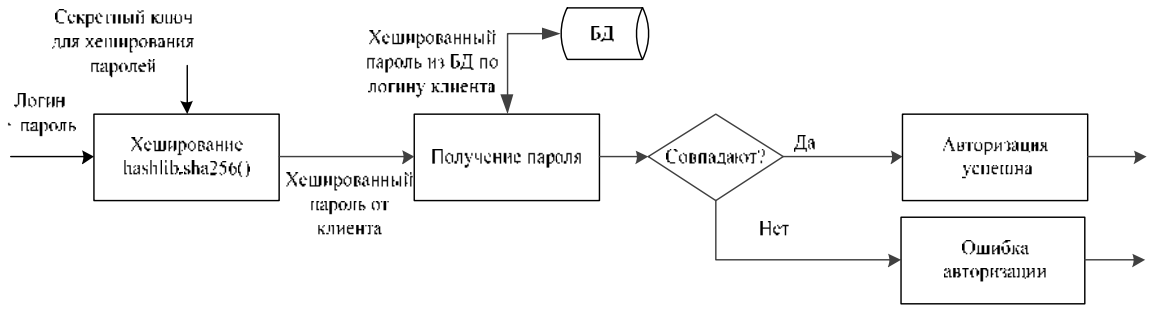


Рис. 2. Процесс авторизации пользователя на сайте

При авторизации пользователь вводит логин и пароль. Пара значений передается на сервер, где пароль проходит процесс валидации. При обращении в базу данных, сервер возвращает пароль согласно введенному логину. Пароль, переданный от клиента хешируется алгоритмом SHA-256 с помощью секретного ключа. В случае если пароль совпал, то авторизация прошла успешно, если нет, пользователю возвращается соответствующее сообщение.

Когда авторизация прошла, возникает потребность в хранении данных авторизации. Для этого можно обратиться к cookies сайта и записать данные сюда. Алгоритм записи данных в cookies представлен на рис. 3.

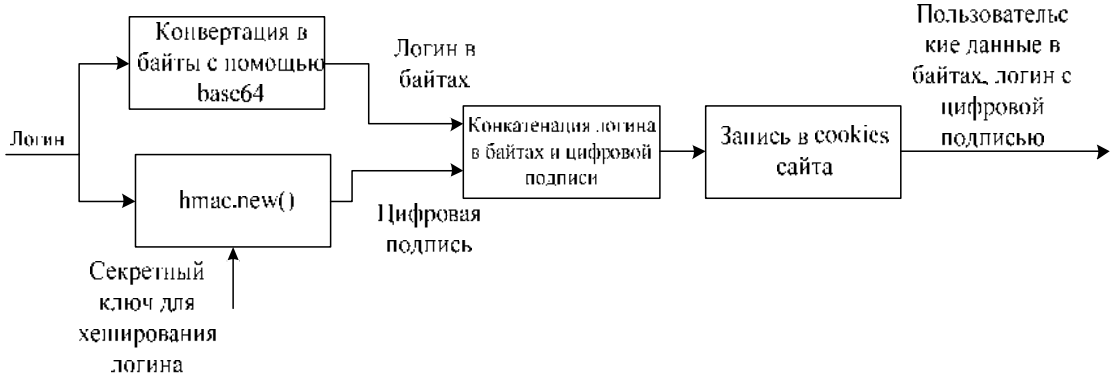


Рис. 3. Алгоритм реализации цифровой подписи логина

Данные в cookies нельзя хранить в открытом виде, поскольку злоумышленник может ими воспользоваться. Для безопасной записи данных в cookies можно применить цифровую подпись. Когда авторизация на сервере подтвердилась, перед тем как записать данные в cookies, логин преобразуется с помощью библиотеки base64 в шестнадцатеричный код. Затем с помощью функции hmac.new() генерируется цифровая подпись логина. В функцию hmac.new() передается логин, секретный ключ для подписи логина

(отличается от секретного ключа пароля), и алгоритм хеширования. Функция hmac.new() в Python представляет собой инструмент для создания объекта HMAC (Hash-based Message Authentication Code). HMAC – это алгоритм аутентификации сообщений, который использует хеш-функцию в сочетании с секретным ключом для генерации кода аутентификации сообщения. Функция hmac.new() принимает два обязательных аргумента: ключ и сообщение. Она возвращает объект HMAC, который может быть использован для генерации кода аутентификации сообщения. Функция hmac.new() в Python является мощным инструментом для генерации кода аутентификации сообщений с использованием HMAC. Она позволяет указать ключ, сообщение и хеш-функцию для создания безопасного кода аутентификации сообщения.

После создания подписи формируются данные для записи в cookies, все пользовательские данные преобразовываются в шестнадцатеричный код, а логин передается в виде конкатенации закодированного логина и его цифровой подписи.

Таким образом, хранение логина в cookies в виде хеша, сгенерированной подписью, позволяет обеспечить целостность и безопасность данных, представленных в открытом виде. Также, наличие цифровой подписи обеспечивает безопасность на протяжении всего времени работы пользователя на сайте за счет постоянной аутентификации подписи.

При каждом переходе на новую страницу cookies проверяются на достоверность. Текущие данные cookies отправляются на сервер. Логин из cookies декодируется, с помощью секретного ключа и хеширования формируется новая электронная подпись. Если полученная таким образом подпись не совпадает с подписью, хранящейся в cookies, то сессия сбрасывается и cookies удаляются, пользователю высвечивается сообщение о попытке взлома.

Таким образом, использование стандартных средств языка программирования Python, а именно: модуль hmac, библиотеку hashlib и библиотеку base64, позволит избежать утечки паролей пользователей, обеспечит возможность безопасной работы на протяжении всей сессии, а также гарантирует целостность пользовательских данных при работе с веб-сайтом. Важно отметить, что используемые в статье библиотеки являются стандартными и не требуют дополнительных этапов настройки и установки.

Список литературы

1. Актуальные киберугрозы: итоги 2022 года // Сайт компании Positive Technologies. URL: https://www.ptsecurity.com/ru-ru/research/analytics/cybersecurity-threatscape-2022/#id9 (дата обращения: 03.09.2023).

2. Симонов Д. Д., Исмоилов М. И., Макаренко Л. Ф. Сравнение эффективности использования технологий, построенных на Python, js и PHP в веб-разработке // Интегрированные автоматизированные системы управления в отраслях транспортно-дорожного комплекса : материалы 80-й науч.-метод. и науч.-исслед. конф. МАДИ (Москва, 24–28 января 2022 г.). М. : Техполиграфцентр, 2022. С. 93–107.

3. Султангулов А. Р., Хисаметдинов Ф. З. Кибербезопасность: хеширование паролей, добавление солей и вычисление скорости работы программы // Неделя науки и технологий : материалы Всерос. науч.-практ. конф. с междунар. участием (Сибай, 12–16 апреля 2021 г.). Сибай : Сибайский информационный центр-филиал ГУПРБ Издательский дом «Республика Башкортостан», 2021. С. 276–277.

4. Головченко Н. А., Печкуров М. А. Обеспечение информационной безопасности данных с помощью библиотек Python // Состояние и перспективы развития современной науки по направлению «Информационная безопасность» : сб. ст. II Всерос. науч.-техн. конф. (Анапа, 19–20 марта 2020 г.). Анапа : Военный инновационный технополис «ЭРА», 2020. Т. 1. С. 369–373.

5. Митричев Д. В., Абросимова Е. А. Безопасность в Python: аспекты защиты авторизации пользователей // Вестник по безопасности : материалы Всерос. науч.-практ. конф. по безопасности (Тольятти, 20–21 декабря 2020 г.). Тольятти : Волжский университет имени В. Н. Татищева (институт), 2020. Вып. 13. С. 46–49.

УДК 621.391:004.056

ОПЫТ ДОРАБОТКИ ПРОГРАММНО-АППАРАТНОГО КОМПЛЕКСА «АРМ "СПЕКТР"» ПОД НОВЫЕ ТРЕБОВАНИЯ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

Д. Ю. Астахов[1], И. Г. Сергина[2], И. Г. Турыгин[3]

[1,2,3] Научно-производственное предприятие «Рубин», г. Пенза

Аннотация. Программно-аппаратный комплекс «АРМ "Спектр"» (далее – изделие) является совокупностью аппаратных и программных средств и предназначен для организации автоматизированного защищенного обмена документальной и речевой информацией между типовыми (аналогичными) АРМ по IP-сетям, с использованием устройства криптографической защиты данных. В статье рассматривается процесс доработки ранее выпущенных изделий в соответствии с новыми требованиями безопасности ФСТЭК России, подробно представлен процесс доработки его составной части – АРМ обмена информации с вычислителем с целью уменьшения уровня ПЭМИН, представлены результаты измерений комплекса на основе сканера для корпусированных изделий Detectus AB RSE 644, представлены результаты доработки вычислителя и дисплея.

Ключевые слова: ПЭМИН, информационная безопасность, компьютер, LVDS, электромагнитные излучения

EXPERIENCE IN FINALIZING THE HARDWARE AND SOFTWARE COMPLEX «ARM "SPEKTR"» PRODUCT FOR NEW INFORMATION SECURITY REQUIREMENTS

D. Yu. Astakhov[1], I. G. Sergina[2], I. G. Turygin[3]

[1,2,3] Scientific-industrial Enterprise «Rubin», Penza

Abstract. The workstation «ARM "SPEKTR"» (hereinafter referred to as the product) is a combination of hardware and software and is designed to organize an automated secure exchange of documentary and speech information between typical (similar) workstations over IP networks using a cryptographic data protection device. The article discusses the process of refining previously released products in accordance with the new safety requirements of the FSTEC of Russia, presents in detail the process of refining its component part - the automated workplace for exchanging information with a computer in order to reduce the TEMPEST level, presents the measurement results of the complex based on a scanner for packaged products Detectus AB RSE 644 , the results of the refinement of the computer and the display are presented.

Keywords: TEMPEST, Information security, computer, LVDS, electromagnetic radiation

Программно-аппаратный комплекс «АРМ «Спектр» (далее изделие) является совокупностью аппаратных и программных средств и предназначен для организации автоматизированного защищенного обмена документальной и речевой информацией между типовыми (аналогичными) АРМ по IP-сетям, с использованием устройства криптографической защиты данных.

В качестве системы защиты информации от несанкционированного доступа используется «Dallas Lock 8.0-C», который ведет полный аудит системы, позволяет настроить мандатный доступ, исключает возможность подключения не учтённых устройств и носителей информации, проводит аутентификацию зарегистрированных пользователей. Для дополнительной защиты при аутентификации используется аппаратный идентификатор.

Изделие позволяет обмениваться информацией различной степени конфиденциальности, включая информацию, содержащую сведения, составляющие государственную тайну. Конструктивно в полном варианте поставки изделие представляет собой пять ударопрочных герметичных кейсов: АРМ обмена информации; Блок защиты информации; Блок подготовки документов; Комплект оборудования специальной телефонной связи (2 шт.). Было отгружено более 30 изделий. С 01.12.2017 вступил в силу приказ № 025 о требованиях по технической защите информации, содержащей сведения, составляющую государственную тайну, утвержденный ФСТЭК России [1]. Возникла необходимость выполнить доработку отправленных изделий под новые требования информационной безопасности. Был проведен ряд конструктивных изменений:

– все внешние разъемы были заменены на военные герметичные;

– добавлены приспособления для опечатывания всех информационных разъемов;

– добавлены заглушки на все информационные разъемы.

В соответствии с новым приказом ужесточались требования для изделия по побочным электромагнитным излучениям и наводкам (ПЭМИН), поэтому возникала необходимость проведения ряда доработок и изменений в части электрических соединений и узлов. С точки зрения информационной безопасности самым проблемным местом был вычислитель и монитор. С помощью

комплекса на основе сканера для корпусированных изделий Detectus AB RSE 644 были проведены исследования на собранном стенде вычислителя и монитора. Результаты измерений представлены на рис. 1–3.

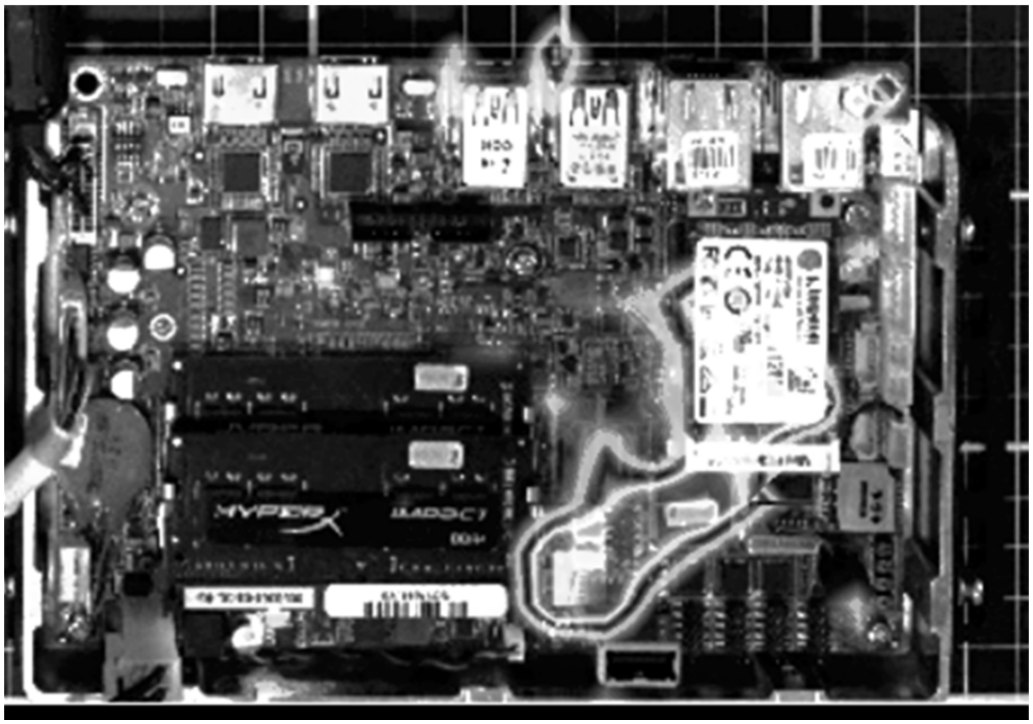


Рис. 1. Результаты исследования ПЭМИН вычислителя

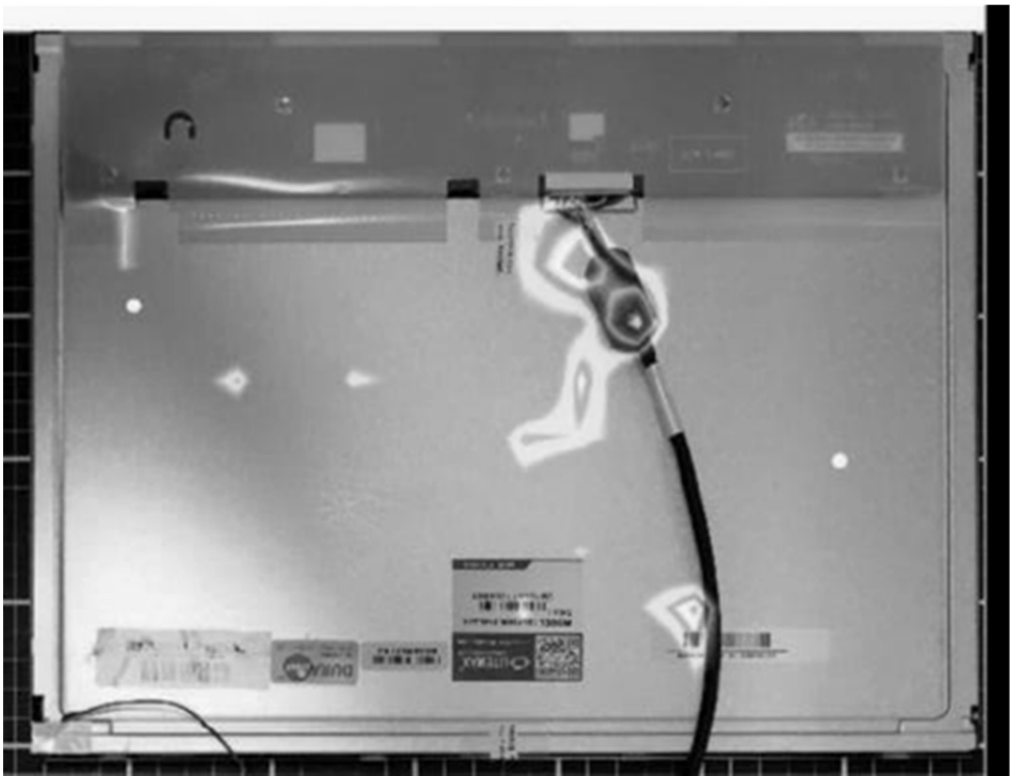


Рис. 2. Результаты исследования ПЭМИН дисплея

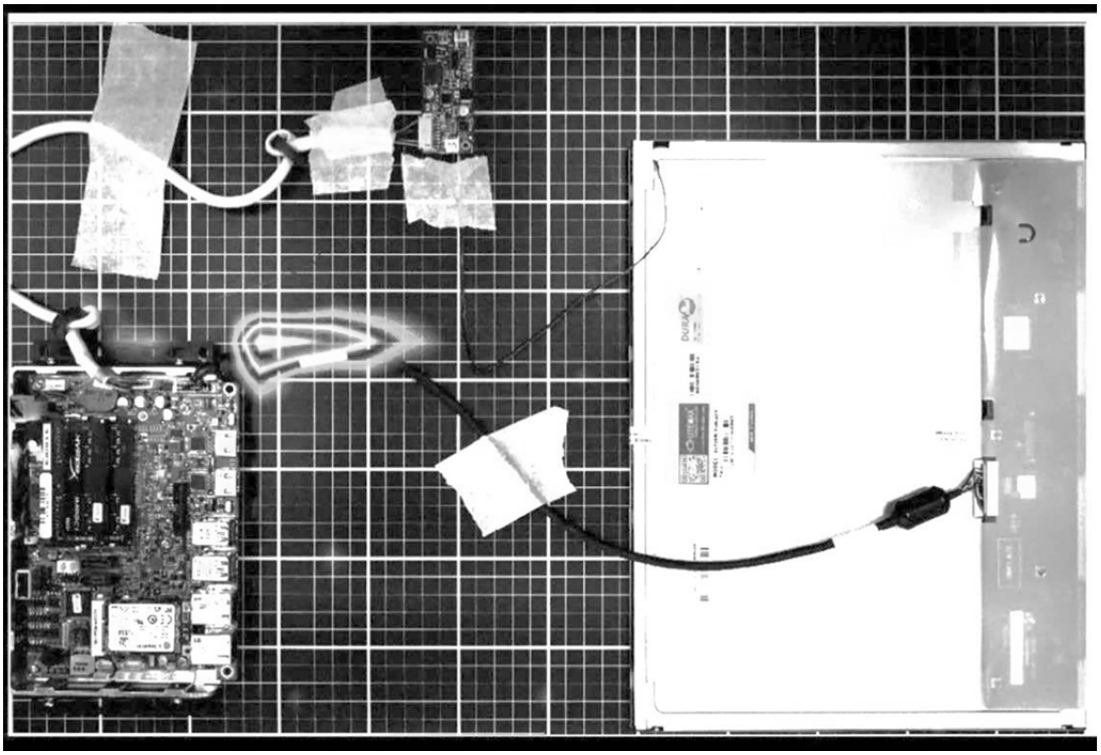


Рис. 3. Результаты измерений ПЭМИН кабеля

На результатах измерений видны области повышенного излучения. Для минимизации излучения были выполнены следующие доработки:

– вычислитель был помещен в металлический кожух;

– LVDS кабель и кабель для драйвера подсветки были переделаны с двойным экранированием;

– с внутренней стороны пластикового кейса АРМ обработки информации было нанесено токопроводящее медное покрытие;

– ЖК-матрица и корпус WEB-камеры были обклеены токопроводящей тканью.

После полной сборки изделия и проведения полного цикла специальных исследований было выявлено, что в АРМ обмена информации принятых решений для уменьшения ПЭМИН недостаточно и требуется дальнейшая доработка. По частотам, уровень которых превышал допустимый уровень, было установлено, что источником излучения является LVDS интерфейс. На плате вычислителя на микросхемы LVDS интерфейса был наклеен радиопоглотитель, так же он был наклеен на всю крышку кожуха вычислителя. Кабель LVDS был дополнительно обмотан токопроводящей тканью на клеевой основе. Данной тканью были также

заклеены все места возможного выхода излучения. На рис. 4–5 показаны результаты доработки.

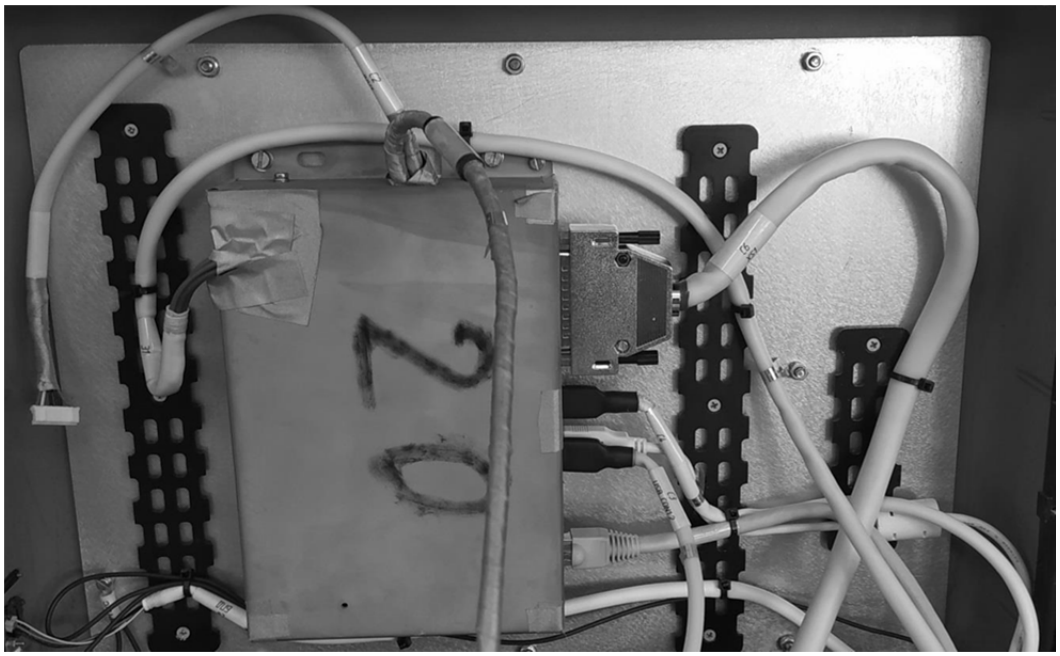


Рис. 4. Результат доработки вычислителя
и кабелей дисплея

Рис. 5. Результат доработки дисплея
и информационных кабелей

Повторное проведение цикла специальных исследований изделия превышение уровня ПЭМИН не выявило.

В ходе доработки изделия было установлено, что для уменьшения уровня ПЭМИН применение общего подхода с добавлением экранов на электронные узлы без применения специализированных материалов недостаточно. На уровень ПЭМИН влияет не только закрытие всех мест возможных утечек специализированными материалами, но и то, как уложены кабели внутри изделия.

Список литературы

1. Требования по технической защите информации, содержащей сведения, составляющую государственную тайну : приказ ФСТЭК России № 025 от 20.10.2016.

УДК 004.056.53

МЕТОДЫ ВЗЛОМА ПАРОЛЕЙ И МЕРЫ ПО ПРОТИВОДЕЙСТВИЮ ВЗЛОМУ ПАРОЛЕЙ

О. В. Зверев[1], А. М. Киселев[2], С. Н. Пелёвин[3], И. С. Ильина[4]

[1,2,3,4] *Научно-производственное предприятие «Рубин», г. Пенза*

Аннотация. Со стремительным развитием и внедрением IT-технологий во все сферы жизни особо остро стоит вопрос о защите данных и, как следствие, идентификации аутентификации пользователей в системе. По мере совершенствования IT-технологий не стояли на месте и методы взлома паролей и велась активная работа по недопущению их взлома. В статье рассмотрены методы взлома паролей и меры по их недопущению, разбитые на два этапа: разработка пароля и этап после разработки пароля. Целью работы является предоставление знаний специалистам информационной безопасности и обычным пользователям по взлому паролей и защиты от него.

Ключевые слова: компьютерная безопасность, аутентификация, взлом паролей, криптоанализ, контрмеры

METHODS OF PASSWORD CRACKING AND MEASURES TO COUNTERACT PASSWORD HACKING

O. V. Zverev[1], A. M. Kiselev[2], S. N. Pelevin[3], I. S. Ilina[4]

[1,2,3,4] *Scientific-industrial Enterprise «Rubin», Penza*

Abstract. With the rapid development and implementation of IT technologies in all spheres of life, the issue of data protection and, as a consequence, the identification of user authentication in the system is particularly acute. With the improvement of IT technologies, the methods of cracking passwords did not stand still, and as a result, active work was carried out to prevent their hacking. The article realises methods of password cracking and measures to prevent them, divided into two stages: password development and the post-password development stage. The purpose of the work is to provide knowledge to information security specialists and ordinary users on password cracking and protection against it.

Keywords: computer security, user authentication, password cracking, cryptanalysis, countermeasures

Введение

Пароль применялся для шифрования информации еще со времен Древнего мира, это привело к образованию такой дисциплины как криптография. В современном мире высоких технологий данная область не только не утратила своей значимости, но и в настоящее время приобрело особую актуальность для аутентификации пользователя в сети Интернет.

RFC 2828 дает определение аутентификации пользователя как процессу проверки подлинности, заявленной системным объектом или для него. Служба аутентификации должна гарантировать, что при подключении не вмешивается третья сторона, маскирующаяся под одну из двух сторон, что обычно касается двух подходов к аутентификации источника данных и аутентификации равноправных пользователей. Также выделяют аутентификацию источника данных, которая обеспечивает подтверждение блока данных, но не обеспечивает защиту от дублирования или модификации данных. Она используется в электронной почте и аутентификация равноправного пользователя, которая обеспечивает подтверждение личности в корпоративных сетях.

Обычно существует четыре способа аутентификации личности пользователя, основанных на: чем-то, что человек знает (например, пароль, PIN-код, ответы на заранее подготовленные вопросы), чем-то, чем он обладает (токен, например, смарт-карта, электронный ключ-карта, физический ключ), чем он является (статические биометрические данные, например, отпечаток пальца, сетчатка глаза, лицо), и что-то, что делает человек (динамические биометрические данные, например, голосовой рисунок, почерк, ритм набора текста) [5][6][4]. В различных системах аутентификация по паролю (то, что известно пользователю) является широко используемой линией защиты от злоумышленников [7]. В схеме аутентификации по паролю идентификатор пользователя определяет, что пользователь авторизован для доступа к системе, и его привилегии. Он также иногда используется для дискреционного контроля доступа, что означает, что другие могут войти в систему, используя ваши личные данные. Когда пользователь вводит имя/логин и пароль, система сравнивает пароль с тем, который сохранен для указанного логина. Однако некоторые пользователи не обращают внимания на конфиденциальность или сложность своих паролей, думая, что у них нет никаких личных файлов в Интернете. Это позволяет взломщикам создавать ущерб всей системе, если

они создают точку входа в систему [8]. Кроме того, быстрые современные процессоры сделали угрозы взлома системы, кражи и повреждения данных проще, чем раньше [9].

История исследования безопасности паролей

Моррис и Томпсон [10] проводят краткое тематическое исследование безопасности паролей и указывают, что пароли с солью (12-битная случайная величина) в это время менее предсказуемы. Джобуш и Олдехофт [11] [12] показывают, что механизмы паролей остаются преобладающим методом идентификации пользователей компьютерной системы. Сначала они дают обзор аутентификации, обсуждают уязвимости механизмов паролей на тот момент, а затем описывают расширение системы паролей UNIX с разрешением использования парольных фраз. Спаффорд [13] [14] проектирует достаточно сложный словарь для проверки паролей, позволяющий предотвратить неправильный выбор пароля с помощью метода, рационального метода.

Казье и Медлин [15] представляют исследование безопасности паролей и времени взлома паролей, сосредоточив внимание на паролях электронной коммерции. Их эмпирическое исследование показывает, что пользователи деловой переписки обычно создают пароли короткие или простые пароли таким образом, чтобы их можно было легко запомнить.

Снайдер [17] разрабатывает шаблон для создания индивидуальных учебных упражнений в области обучения безопасности со взломом/восстановлением паролей, реализованных на практике. Фернелл [18] обсуждает вопросы защиты веб-сайта паролем в трех аспектах, включая то, предоставляют ли сайты какие-либо рекомендации при выборе пользователями пароли, налагаются ли на пароли какие-либо ограничения и предлагаются ли средства для оказания помощи пользователям, которые забыли пароли. Плага [19] проводит исследование подходящих вариантов использования биометрических ключей. Виндер [20] кратко представляет некоторые из наиболее часто используемых методов взлома паролей, включая атаку по словарю, атаку методом перебора, атаку по радужной таблице, вредоносное ПО, автономный взлом, угадывание и т.д. Другие связанные с этим работы включают в себя то, что Бонно [21] проводит анализ списка из 70 миллионов паролей.

Использование пароля при удаленной аутентификации пользователя

При удаленной аутентификации пользователя используется протокол паролей, когда пароль используется для аутентификации [2] [4]. В схеме протокола паролей пользователь сначала передает идентификатор удаленному хосту; затем хост генерирует случайное число (одноразовый номер), и случайное число возвращается пользователю, в то же время хост сохраняет хэш-код пароля; пользователь хэширует пароль случайным числом и отправляет ее хосту; хост сравнивает сохраненную информацию с полученной от пользователя, чтобы проверить, соответствуют ли она системе или нет.

1. Традиционные методы взлома паролей

Поскольку пароли остаются наиболее широко используемым механизмом аутентификации пользователей, получение паролей по-прежнему является распространенным и эффективным методом атаки [26].

Традиционные методы взлома паролей включают в себя кражу, обман, анализ пользователей, анализ алгоритмов и полное угадывание и т.д. который будет представлен ниже.

1. Кража

Кража паролей может быть осуществлена путем осмотра рабочего стола пользователя, подглядывание, прослушивания подключения к сети с целью получения незашифрованных паролей, получения доступ к базе паролей и вредоносным программам. При краже хакеры будут маскироваться под курьера посылок, специалиста по обслуживанию или кого-то еще, чтобы получить доступ к офису. После того, как они попадают в здание, униформа обслуживающего персонала предоставляет своего рода бесплатный пропуск, позволяющий им бродить по зданию и записывать пароли, вводимые настоящими сотрудниками [20]. При атаке вредоносным ПО обычно устанавливается регистратор ключей или средство очистки экрана, которое может записывать все, что вводит пользователь, и делать снимки экрана во время процесса входа в систему. Кроме того, некоторые вредоносные программы пытаются проверить наличие файла паролей клиента веб-браузера и скопировать доступные пароли из истории посещенных страниц.

2. Обман

Другой способ получить пароль – это обмануть пользователей с помощью социальной инженерии или фишинга в режиме онлайн. Один из случаев – позвонить в офис, выдавая себя за профессионального специалиста по ИТ-безопасности, и запросить пароли доступа к учетной записи пользователя (или сети) [15]. В результате фишинговой атаки пользователи получат фишинговое электронное письмо, содержащее ссылки, ведущие на поддельные веб-сайты, например, онлайн-банкинг и платежи и т.д., Что создает угрозу безопасности учетных записей.

3. Анализ пользователей

Пользователи, как правило, генерируют пароли на основе того, о чем они обычно общаются в социальных сетях или где-либо еще хакеры, скорее всего, изучат такого рода информацию и сделают несколько предположений во время взлома паролей. Хакеры могут сократить время взлома паролей, проанализировав специальных пользователей в соответствии с их характеристиками, такими как имя, должность, интересы, семьи, хобби и так далее. Один из таких видов атак известен как «Паучья атака». Хакеры понимают, что многие корпоративные пароли генерируются путем подключения к самому бизнесу. Они пытаются составить пользовательские списки слов, изучая корпоративную литературу, материалы веб-сайта, список клиентов и т.д.

4. Анализ алгоритма

Атаки с анализом алгоритмов фокусируются на используемых алгоритмах шифрования, таких как криптоаналитические атаки, которые также используются при расшифровке зашифрованного текста [9]. Это зависит от алгоритма, некоторого знания общих характеристик текста и некоторой выборки пар открытого текста с зашифрованным текстом. Этот вид атаки использует характеристики алгоритмов, чтобы попытаться вывести определенный открытый текст или ключи.

5. Полное угадывание

Широко используемым видом методов взлома паролей является полное угадывание, включающее атаку по словарю, атаку методом перебора, гибрид словаря и brute-force [15], атаки с использованием радужных таблиц и т.д.

При атаке по словарю хакеры, пытающиеся получить доступ к компьютеру или сети пользователей, используют большой словарь, содержащий возможные пароли (часто используемые пароли, например, общеупотребительные словарные слова, сочетание нескольких слов, игру слов и т.д.).

Общий подход заключается в применении одного и того же метода шифрования к словарю паролей для сравнения с копией зашифрованного файла, содержащего пароли. Если в методе шифрования использовалась хэш-функция, то каждый пароль по словарю должен быть хэширован с использованием каждого значения соли для сравнения с сохраненными хэш-значениями. Бергер и др. [28] также представляют атаки по словарю с использованием акустических излучений клавиатуры.

Атака методом перебора угадывает пароль, используя случайный подход, пробуя разные пароли и надеясь, что один из них сработает [29]. В случае со словарем этот подход отличается атакой с использованием слов, не входящих в словарь, которые могут содержать все возможные буквенно-цифровые и даже специальные комбинации символов, такие как «aaaaa000», «zzzzz9999» и «bbb&&99$00». Если при атаке методом перебора используется некоторый логический анализ, время взлома пароля иногда может быть сокращено, например, за счет использования ранее упомянутого пользовательского анализа (с использованием имен пользователей, информации о работе и хобби). Предполагая, что пароли пользователей содержат более нескольких символов, атака методом перебора может в конечном итоге взломать пароли, даже если на это уйдет много времени.

Использование моделей распределенных вычислений и ботнетов может сократить время взлома. Атака методом перебора также используется при расшифровке зашифрованного текста, когда она пробует все возможные ключи в зашифрованном тексте до тех пор, пока не будет получен понятный перевод открытого текста.

В атаках с использованием радужной таблицы хакеры используют радужную таблицу, которая представляет собой список предварительно вычисленных хэш-значений для всех зашифрованные пароли со всеми солями [30]. Время, затрачиваемое атакой rainbow table на взлом пароля, сокращается до времени, необходимого для поиска его в списке. Такого рода атаке можно было бы противостоять, используя достаточно большое значение соли и достаточно большую длину хэша.

Если взлом пароля осуществляется онлайн, система защиты обычно блокирует пользователей после того, как они не смогли войти в систему более трех раз, чтобы заблокировать автоматизированное программное обеспечение для угадывания. Однако оффлайновые атаки обычно избегают этого. При атаке в автономном режиме хакеры сначала взламывают систему, чтобы украсть зашифрованные файлы паролей или подслушать зашифрованный обмен данными в Интернете. Тогда взломщику паролей может потребоваться столько времени, сколько потребуется, чтобы попытаться взломать код, не предупреждая целевую систему или отдельного пользователя.

2. Метод, основанный на марковских цепях

Марковские модели используются в некоторых инструментах для взлома паролей или восстановления данных. Цепи Маркова могут сократить пространство поиска, которое потребуется при использовании атаки методом перебора.

Нараянан и Шматиков [33] используют фильтры Маркова для определения фонетического сходства со словами на родных языках пользователей (ищут только «запоминающиеся» строки). И используют конечную автоматизированную модель для работы с неалфавитными символами в паролях. Используемые марковские модели включают модель нулевого порядка и модель первого порядка. В марковской модели нулевого порядка символы независимы друг от друга и генерируются в соответствии с распределением вероятностей по формуле описания (P используется для марковской вероятности, v – частота символа):

$$P(\alpha) = \prod_{X \epsilon \alpha} v(x) \qquad (1)$$

С другой стороны, символы в первом порядке Марковская модели зависимы, и каждому символу будет присвоена вероятность путем просмотра предыдущих символов (символ « | », обозначающий, что правая часть является предпосылкой левой части):

$$P(x_1, x_2, \dots, x_n) = v(x_1) \prod_{i=1}^{n-1} v(x_{i+1} | x_i) \qquad (2)$$

Цепи Маркова также могут быть использованы в качестве мощного инструмента для улучшения взлома распределенных таблиц и радужных таблиц.

3. Вероятностные контекстно-свободные грамматики

PCFG используют вероятностный контекстно-свободный Грамматики для взлома паролей. Сначала они создают вероятностную контекстно-свободную грамматику, основанную на обучающем наборе раскрытых паролей. Они включают информацию о распределении вероятностей пользовательских паролей для генерации шаблонов паролей. Они используют грамматику для генерации правил искажения слов, которые будут использоваться в период подбора пароля при атаке по словарю.

Этот метод показал улучшение более чем четверть по сравнению с общедоступной программой для взлома паролей «John the Ripper», протестированной на множестве раскрытых паролях.

Меры противодействия взлому паролей

Защита паролей от компрометации и несанкционированного использования является важнейшей проблемой, поскольку пароли остаются наиболее популярным способом аутентификации. Как правило, меры противодействия взлому паролей могут быть приняты в два этапа, т.е. на этапе разработки пароля и после генерации.

1. Этап разработки

1. Обучение пользователей

Пользователи могут быть проинформированы о важности использования надежных паролей и обучены тому, как генерировать труднодоступные пароли, используя некоторые стратегии выбора паролей. Например, пароль должен содержать:

– буквы (обе заглавные Буквы и строчные буквы);

– цифры и специальные символы;

– длина пароля не должна быть меньше определенного числа.

2. Динамический пароль

Коул [26] приводит введение одноразового пароля, динамического пароля и статического пароля, из которых одноразовый пароль является одним из способов обеспечить высокий уровень безопасности. В схеме одноразового пароля новый пароль требуется каждый раз, когда пользователь входит в учетную запись, чтобы хакеры не могли использовать предварительно скомпрометированный пароль.

Динамический пароль указывает, что пароль меняется часто или через короткий промежуток времени, в то время как статический пароль означает, что пароль остается неизменным в течение всего времени входа в систему.

Организация может ввести требования к пользователям относительно в периодической смене паролей, например, еженедельно, ежемесячно или каждые полгода.

Продолжительность временного интервала может быть основана на важности защищаемой информации.

3. Использование токена

Пароль может быть сгенерирован с использованием некоторых токенов безопасности, которые используются для облегчения аутентификации авторизованных пользователей компьютерных служб. Токеном может быть физическое устройство, такое как смарт-карты (оно также может относиться к программному токену или виртуальному токену). Пароль, указанный на токене, может регулярно изменяться с определенным интервалом времени, что обеспечивает механизм динамического паролирования и снижает ценность украденных паролей из-за их короткого срока действия. Кроме того, постоянная смена пароля снижает вероятность успешного взлома с помощью атаки методом перебора, если злоумышленник использует список паролей в течение одной смены. Проблема заключается в том, как справиться с синхронностью токена и сервера из-за того, что произойдет временная задержка, прежде чем пароль достигнет токена с сервера. После того, как пользователь введет пароль, указанный на токене, пароль на сайте сервера может быть уже изменен на следующий из-за временной задержки. Токен также может быть оснащен встроенными алгоритмами переключения. В этом случае пользователь вводит цифры, указанные на токене, после чего калькулятор сгенерирует пароль, используя введенный алгоритмы.

4. Сгенерированные компьютером пароли

Пользователи также могут использовать сгенерированный компьютером пароль для своей учетной записи. Используя некоторый предварительные требования, сгенерированный компьютером пароль обычно имеет определенную длину, содержит специальные символы и является непроизносимым, что затрудняет успешный взлом хакерами в течение короткого времени. Однако пользователям нелегко запомнить сгенерированный компьютером пароль из-за того, что он по большей части бессмысленен.

2. После генерации паролей

1. Реактивная проверка пароля

В рамках стратегии реактивной проверки паролей система периодически запускает свой собственный инструмент для взлома

паролей, чтобы найти пароли, которые можно угадать. Система отменит угаданные пароли и уведомит об этом пользователей. Недостатками являются то, что это очень сильно потребляет ресурсы, и хакеры также могут использовать эту стратегию для поиска слабых паролей, если они получат копию файла паролей.

2. Проактивная проверка пароля

Еще одним способом отклонить слабые пароли является проактивная проверка паролей. В отличие от реактивной проверки пароля, проактивная проверка пароля позволяет пользователям выбирать свой собственный пароль. Однако система проверит, является ли пароль допустимым или нет. Цель состоит в том, чтобы пользователи могли выбирать запоминающиеся пароли, которые трудно угадать. Следует проводить проактивную проверку паролей, чтобы предотвратить попадание легко угадываемых паролей в системе в первую очередь в соответствии с существовавшими на тот момент методами взлома паролей, которые включают в себя проверку соответствующей личной информации пользователя (имя, инициалы, название учетной записи и т.д.), проверку слов из различных словарей и различные перестановки слов с заглавными буквами, проверку иностранных языковые слова для иностранных пользователей и пробование пар слов.

Предлагаемое средство проверки паролей обладает некоторыми характеристиками, включая отклонение любого пароля из набора распространенных паролей (например, словарных слов, паролей, основанных на имени пользователя или учетной записи, а также паролей, длина которых меньше определенной), что позволяет различать каждого пользователя и сайт в своих тестах, имея шаблон средство сопоставления, которое может храниться в тестах, легко настраивается и т.д.

3. Шифрование пароля

Защита от шифрования паролем включает в себя хэш-функции и использование солей. В области компьютерной безопасности и криптографии хэш-функции относятся к алгоритмам, которые принимают входные данные переменного размера и возвращают строку фиксированного размера в качестве хэш-значения. Такой подход гарантирует, что любые изменения во входных данных приведут к другому хэш-значению. Существует несколько общих характеристик хэш-функции, включая простоту вычислений (легко вычислить значение хэша для любого заданного сообщения), устойчивость к предварительному проекту пароля (вычислительно невозможно сгенерировать сообщение с заданным

хэшем), устойчивость ко второму предварительному проекту пароля (учитывая входные данные A, трудно найти другой другой вход B, чтобы они имели одинаковое значение хэша) и устойчивость к столкновениям (вычислительно невозможно найти два разных сообщения с одинаковым хэшем) и т.д.

Существует несколько часто используемых хэш-функций, таких как MD5 (выдает 128-битное хэш-значение) и SHA (алгоритм безопасного хэширования, указанный в защищенном хэше. Алгоритмы SHA (0-3) разработаны с учетом характеристик быстрых вычислений и эффективной реализации в аппаратном обеспечении, но эти характеристики также имеют некоторые недостатки, такие как неэффективность при предотвращении взлома паролей, таких как атаки с использованием радужной таблицы. За исключением шифрования паролей, хэш-функция также используется в основном в цифровых подписях.

Соль в криптографии и компьютерной безопасности определяет случайные данные (например, одноразовый номер), которые используются в качестве дополнительных входных данных для односторонней функции, которая хэширует пароль. Новая соль генерируется системой случайным образом для каждого нового пароля. Как правило, соль и пароль объединяются и шифруются с помощью хэш-функции, а затем выходное хэш-значение сохраняется в файле. Использование соли могло бы эффективно замедлить атаки с использованием предварительно вычисленных таблиц rainbow (радужных таблиц), поскольку хакерам приходится хэшировать каждый потенциальный пароль, используя все соли. Соли широко используются в различных компьютерных системах – от системных учетных данных UNIX до интернет-безопасности.

Существует применение двух солей – общедоступной и секретной. Общедоступная соль выполняет ту же функцию, что и соль в это время, в то время как секретная соль генерируется случайным образом при вводе пароля в первый раз и будет отброшена разработанной системой после использования. Пользователям не нужно ничего знать о секретной соли, и секретная соль нигде не хранится.

Результаты показывают, что использование разработанных двух солей без изменения какой-либо другой части механизма шифрования может затруднить взлом пароля в 100–1000 раз.

4. Контроль доступа

Общая мысль заключается в том, что если хакеры не смогут получить файлы с паролями (или зашифрованными паролями),

то эффективность взлома паролей сильно снизится, поскольку хакеры не смогут выполнить автономное угадывание. Согласно этому анализу, контроль доступа к файлам паролей очень важен и эффективен для предотвращения взлома паролей, и этот метод используется в некоторых системах.

Например, современные операционные системы UNIX хранят файл хэшированного пароля в маршруте, доступном только программам, работающим с повышенными привилегиями. Пароль метод контроля доступа к файлам также включает в себя некоторые уязвимости, такие как слабость операционных систем, которые разрешают доступ к файлу паролей, случайное разрешение на чтение файла паролей, перехватывающая передача пароля по сети и слабые схемы вызова или ответа в сетевых протоколах и т.д.

В связи с этим предлагаются требования по проектированию без облачного хранилища для достижения высокого уровня безопасности.

Вывод

Взаимодействие аутентификации, конфиденциальности и целостности лежит в основе важнейших аспектов модели доверия, которые должны применяться для обеспечения безопасности, из которых хорошо известной проблемой является аутентификация как пользователей, так и пакетов данных. Пароль остается наиболее широко используемым методом обеспечения аутентификации. Однако пароли подвержены некоторым уязвимостям, таким как автономная атака, атака на конкретную учетную запись, атака популярными паролями, анализ пользователей, захват рабочей станции, использование ошибок пользователей, электронный мониторинг и т.д. Хакеры обычно используют эти уязвимости паролей позволяют осуществлять их атаку. В этой работе сначала обсуждаются широко используемые методы взлома паролей и классифицируются они на несколько категорий, включая традиционные методы (кража, обман, анализ пользователей, анализ алгоритмов и полное угадывание), а также марковскую модель, вероятностные методы, контекстно-свободная грамматику. Затем представлены контрмеры для взлома паролей в два этапа, включая этап разработки пароля (например, обучение пользователя, динамический пароль, использование токенов и сгенерированный компьютером пароль) и средства защиты после генерации паролей (например, реактивная проверка пароля, проактивная проверка пароля, шифрование пароля и контроль доступа к файлам паролей).

Список литературы

1. Stallings W., Brown L. Computer Security: Principles and Practice, Prentice Hall. 2nd Edition. 2011.

2. Wood H. M. The use of passwords for controlled access to computer resources, NBS Special Publication 500-9, U.S. Dept. of Commerce/NBS, 1977.

3. Dasgupta D., Saha S. Biologically inspired password authentication system. In Proceedings of the 5th Annual Workshop on Cyber Security and Information Intelligence Research: Cyber Security and Information Intelligence Challenges and Strategies (CSIIRW '09), 2009.

4. Seeley D. Password cracking: a game of wits. Found in: Communications of the ACM. 1989. Vol 32, № 6. P. 700–703.

5. Bishop M., Klein D. V. Improving system security via proactive password checking // Computers & Security. 1995. Vol. 14, iss. 3. P. 233–249.

6. Landwehr C. E. Computer security // International Journal of Information Security. 2001. Vol. 1, iss. 1. P. 3–13.

7. Morris R., Thompson K. Password Security: A Case History. Commun. ACM. 1979. Vol. 22, № 11. P. 594–597.

8. Jobusch D. L., Oldehoeft A. E. A survey of password mechanisms: Weaknesses and potential improvements. Part 1 // Computers & Security. 1989. Vol. 8, iss. 7. P. 587–604.

9. Jobusch D. L., Oldehoeft A. E. A survey of password mechanisms: Weaknesses and potential improvements. Part 2 // Computers & Security. 1989. Vol. 8, iss. 8. P. 675–689.

10. Spafford E. H. OPUS: Preventing weak password choices // Computers & Security. 1992. Vol. 11, iss. 3. 1992. P. 273–278.

11. Spafford E. Observations on Reusable Password Choices // Proceedings of the 3rd USENIX Security Workshop. 1992. P. 299–312.

12. Cazier J. A., Medlin D. B. Password security: An empirical investigation into e-commerce passwords and their crack times // EDPACS. 2006. Vol. 15 (6). P. 45–55.

13. Snyder R. Ethical hacking and password cracking: a pattern for individualized security exercises // In Proceedings of the 3rd annual conference on Information security curriculum development (InfoSecCD'06). ACM, New York, NY, USA, 2006. P. 13–18.

14. Furnell S. An assessment of website password practices // Computers & Security. 2007. Vol. 26, iss. 7–8. P. 445–451.

15. Plaga R. Biometric keys: suitable use cases and achievable information content // International Journal of Information Security. 2009. Vol. 8, iss. 6. P. 447–454.

16. Winder D. Top ten password cracking methods. URL: http:// www.pcpro.co.uk/features/371158/top-ten-password-crackingtechniques (retrieved, April 1st, 2013).

17. Bonneau J. The Science of Guessing: Analyzing an Anonymized Corpus of 70 Million Passwords // Security and Privacy, IEEE Symposium on. 2012. P. 538–552.

18. Cole E. Network Security Bible. 2nd / ed. by Ronald Krutz and James W. Conley. Wiley Publishing, Inc. 2009.

19. Berger Y., Wool A., Yeredor A. Dictionary attacks using keyboard acoustic emanations // In Proceedings of the 13th ACM conference on Computer and communications security (CCS '06). New York, NY, USA. 2006. P. 245–254.

20. Kedem G., Ishihara Y. Brute force attack on UNIX passwords with SIMD computer // In Proceedings of the 8th conference on USENIX Security Symposium. Berkeley, CA, USA, 1999. Vol. 8.

21. Theocharoulis K., Papaefstathiou I., Manifavas C. Implementing Rainbow Tables in High-End FPGAs for Super-Fast Password Cracking // International Conference on Field Programmable Logic and Applications. 2010. P. 145–150.

22. Narayanan A., Shmatikov V. Fast Dictionary Attacks on Passwords Using Time-Space Tradeoff, CCS'05, Alexandria, Virginia, 2005.

УДК 004.056.53

ОБЕСПЕЧЕНИЕ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ ЛИНИЙ ЭЛЕКТРОСНАБЖЕНИЯ МОБИЛЬНЫХ ОБЪЕКТОВ, ИСПОЛЬЗУЮЩИХ ВСТРОЕННЫЕ ИСТОЧНИКИ ЭЛЕКТРИЧЕСКОЙ ЭНЕРГИИ

А. А. Кочетков[1], М. А. Борисов[2], И. В. Хохлов[3], П. П. Первушкин[4]

[1,2,3,4] *Научно-производственное предприятие «Рубин», г. Пенза*

Аннотация. Целью работы является определение возможности уменьшения применения технических средств защиты информации в составе подвижных объектов при работе от встроенных источников электрической энергии без использования промышленных источников. Предлагается комбинированный вариант электроснабжения с использованием электроагрегатов и солнечной электростанции. Проводится расчет вырабатываемой энергии.

Ключевые слова: мобильные объекты, информационная безопасность, система электроснабжения, встроенные источники электрической энергии

ENSURING INFORMATION SECURITY OF POWER SUPPLY LINES OF MOBILE OBJECT, USING BUILDING SOURCE OF ELECTRIC ENERGY

A. A. Kochetkov[1], M. A. Borisov[2], I. V. Khokhlov[3], P. P. Pervushkin[4]

[1,2,3,4] *Scientific-industrial Enterprise «Rubin», Penza*

Abstract. The aim of the work is to determine the possibility of reducing the use of technical means of information protection as part of moving objects when operating from built-in sources of electrical energy without the use of industrial sources. A combined variant of power supply using electric units and a solar power plant is proposed. The generated energy is calculated.

Keywords: mobile objects, information security, power supply system, built-in sources of electrical energy

В связи с активным развитием машиностроительной отрасли широкое распространение получают высокомобильные объекты специального назначения, к которым относятся мобильные офисы, ситуационные центры, военная и специальная техника и т.п.

Подобные объекты строятся как на основе шасси транспортных средств (микроавтобусов, автобусов, грузовиков), так и на основе быстровозводимых помещений (контейнерного типа, мобильные пневмокаркасные палатки и т.д.). Пример мобильного объекта на шасси автомобиля КАМАЗ представлен на рис. 1.

Рис. 1. Внешний вид мобильного объекта
на шасси автомобиля КАМАЗ

Любой из подобных объектов несет в своем составе информационную систему (подсистему) для обработки, передачи или хранения информации, информация может быть различного содержания, конфиденциальности, степени ценности. Ценность информации – ее свойство, характеризующее пригодность к практическому использованию в различных областях целенаправленной деятельности человека [1].

В соответствии с требованиями федеральных законов информация ограниченного доступа подлежит защите. Защита информации осуществляется путем принятия правовых, организационных и технических мер, неправомерного воздействия на информацию (уничтожения, модифицирования, неправомерного блокирования доступа к информации) [2].

Наряду с информационной системой в состав подобного рода объектов входит и система электроснабжения для обеспечения

потребителей основных технических средств (вычислительная техника, оборудование связи и телекоммуникации) и вспомогательных технических средств (освещение, климатическая техника и т.д.).

Базовая система электроснабжения состоит:

– из источников электрической энергии;

– преобразователей электрической энергии;

– распределительных устройств;

– заземляющих устройств;

– кабельной продукции.

Дополнительно стоит отметить, что большинство технических средств имеют в своем составе встроенные преобразователи электрической энергии и имеют встроенную клемму для подключения к заземлению объекта.

Основные технические средства взаимодействуют и оказывают влияние на вспомогательные технические средства и энергетические системы, образуя тем самым угрозу утечки информации по цепям электроснабжения и заземления.

Традиционными способами защиты от подобного рода угроз являются:

– использования технических средств для организации гальванических развязок цепей электроснабжения объекта от общей энергосистемы;

– использование технических средств для электромагнитного зашумления отходящих линий;

– экранирование как кабельных линий, так и шкафов для размещения оборудования;

– использование помехоподавляющих фильтров;

– создание контролируемых зон для размещения собственных источников электрической энергии (электроагрегатов) и устройств заземления;

– организационные меры.

Все эти меры направлены на защиту от неправомерного получения информации с использованием линий от вспомогательных технических средств, выходящих за пределы контролируемой зоны.

В данной работе предлагается рассмотреть возможность минимизировать количество отходящих линий электроснабжения от мобильных объектов. Для этого предлагается рассмотреть возможность организации частичного или полного электроснабжения объектов от встроенных источников электрической энергии. Это

позволит уменьшить количество оборудования и кабельного имущества для обеспечения защиты информации по указанным техническим каналам.

А в связи с тем, что рассматриваемые в этой работе объекты являются высокомобильными и зачастую место их функционирования удалено от промышленных источников электроэнергии, это позволит попутно решить проблемы стабильного электроснабжения.

Самым распространенным встроенным источником электрической энергии для таких объектов является электроагрегат функционирующий от двигателя внутреннего сгорания и накопители электрической энергии (аккумуляторные батареи).

При длительной работе электроагрегатов необходимо уделять внимание своевременному проведению технического обслуживания, ремонта и обеспечению ГСМ. Электроагрегаты не все время своей работы функционируют на номинальной мощности, а зачастую работают в режиме малых нагрузок, что ведет к снижению КПД установки, расходу ГСМ, наработке моточасов и, как следствие, ускоренному выходу из строя.

Таким образом использование электроагрегатов, работающих от ДВС, как источников электрической энергии не всегда эффективно. Для повышения эффективности работы систем электроснабжения без создания дополнительных линий электроснабжения, требующих защиты, возможно рассмотреть установку дополнительных источников электрической энергии. Таким источником может стать фотоэлектрический преобразователь. Разместить его возможно на любых горизонтальных или на наклоненных под небольшим углом поверхностях. Пример размещения представлен на рис. 2.

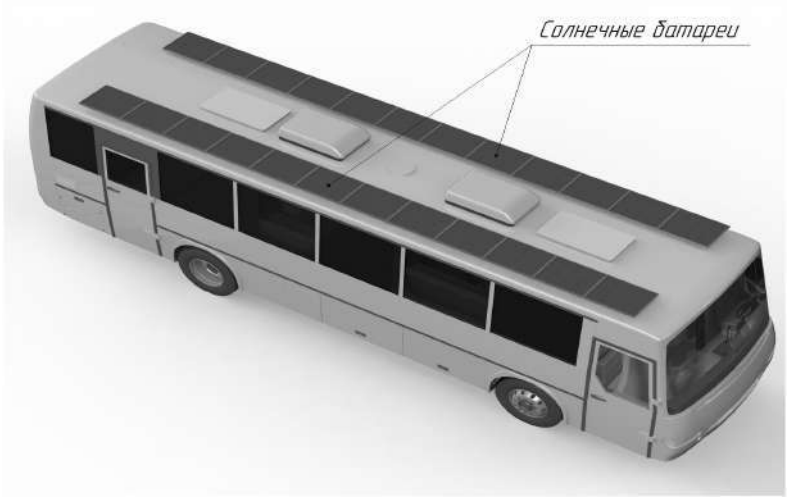


Рис. 2. Размещение солнечных панелей

Электрическая энергия, вырабатываемая солнечными батареями, зависит от различных факторов: времени года, географического положения и угла наклона солнечной батареи. Эти факторы необходимо учитывать при расчете мощности солнечных панелей. Также расчет мощности солнечных батарей необходимо проводить для конкретного региона, проанализировав при этом статистические данные о солнечной активности за несколько лет. Под солнечной активностью понимается среднегодовая инсоляция. Инсоляция – облучение поверхностей солнечным светом (солнечной радиацией), поток солнечной радиации на поверхность [3].

Отечественные наработки в сфере применения альтернативных источников электрической энергии в подвижных объектах достаточно малы. В связи с этим предлагается один из вариантов применения солнечных панелей (батарей) в подвижном объекте. В качестве примера предлагается вариант размещения солнечных батарей SDM-50 на крыше автобуса, а также проводится расчет вырабатываемой электрической энергии.

Основные характеристики солнечной батареи приведены в табл. 1 [4].

Таблица 1

Модель	SDM-50
Номинальная мощность, Вт	50
Рабочее напряжение, В	17,6
Рабочий ток, А	2,85
КПД модуля	13,19
Диапазон рабочих температур, °С	–40 – +85
Габаритные размеры, мм	670×540×25
Масса, кг	4
Класс защиты (IP)	67

Расчет вырабатываемой энергии солнечными батареями, расположенными на автобусе, проводится для Московского региона. Энергия, вырабатываемая солнечными батареями, рассчитывается по формуле:

$$W = \frac{k \times E \times P_w}{P}$$

где W – энергия, вырабатываемая солнечными панелями, Вт*ч; E – среднесуточная солнечная инсоляция, кВт·ч/м^2; P_w – мощность солнечных панелей, Вт; P – пиковая плотность потока солнечной энергии (1000 Вт/м^2) [6]; k – коэффициент, характеризующий

потерю мощности солнечных элементов при нагреве на солнце и её дальнейшем преобразовании (обычно принимают равным 0,5–0,7).

Отношение E/P определяет «так называемое количество пикочасов, т.е. условное время, в течение которого солнце светит как бы с интенсивностью 1000 Вт/м^2» [5].

Среднегодовая и среднесуточная солнечная инсоляция приведена в табл. 2 [6].

Таблица 2

Москва, широта 55.7	Среднегодовая инсоляция кВт·ч/м^2	Среднесуточная инсоляция кВт·ч/м^2
Горизонтальная панель	1020.7	2,8
Вертикальная панель	908.3	2,49
Наклон панели – 45.0°	1173.7	3,22
Вращение вокруг полярной оси	1514.3	4,15

Подставляя данные в вышеприведенную формулу, получаем:

$$W = \frac{0{,}6 \times 2{,}8 \times 1200}{1000} \approx 2 \text{ кВт} \cdot \text{ч}$$

Энергия $W \approx 2$ кВт·ч (≈7,2 МДж) вырабатывается солнечными батареями за сутки.

Вся вырабатываемая энергия накапливается в аккумуляторных батареях и в дальнейшем может использоваться потребителями.

Для примера рассчитаем потребляемую мощность осветительного оборудования. Если для освещения используются автотранспортные светодиодные светильники (например, светильник автотранспортный СИЕУ. 453754.009). Мощность одного светильника составляет 5 Вт. За сутки такому светильнику необходимо 120 Вт. Отсюда следует, что вырабатываемая солнечными батареями энергия достаточна для работы осветительного оборудования внутри автобуса, состоящего из десяти светильников.

В заключении стоит отметить, что использование встроенных источников электрической энергии (совокупно электроагрегатов, солнечных панелей и накопителей электрической энергии) позволяет исключить технические средства защиты информации от утечки по цепям электроснабжения, отходящим от объектов, но не исключает взаимного влияния на основные технические

средства побочных электромагнитных полей от вспомогательного оборудования.

Список литературы

1. Першиков В. И., Савинков В. М. Толковый словарь по информатике. М. : Финансы и статистика, 1991. 543 с.

2. ГОСТ Р 50922–2006. Защита информации. Основные термины и определения. Введ. 2008-02-01. М. : Стандартинформ, 2007. 12 с.

3. Инсоляция. URL: https://altenergiya.ru/sun/chto-takoe-insolyaciya.html

4. Солнечная батарея SDM-50 монокристаллическая. URL: https://al-energy.ru/magazin/product/solnechnaya-batareya-sdm-50-monokristallicheskaya

5. Охоткин Г. П. Методика расчета мощности солнечных электростанций // Вестник Чувашского университета. 2013. Вып. 3.

6. Количество солнечной энергии в регионах. URL: http://realsolar.ru/13890.html

УДК 004.056.52

ПРОБЛЕМЫ ЗАЩИЩЕННОСТИ КАНАЛОВ СВЯЗИ ЦАТС В СЕТЯХ ПЕРЕДАЧИ ДАННЫХ

А. В. Сериков[1], Л. В. Куц[2], Д. М. Зиновьев[3], Р. Ю. Романихин[4], С. А. Гужова[5]

[1,2,3,4,5] *Научно-производственное предприятие «Рубин», г. Пенза*

Аннотация. Рассмотрены основные вопросы по доработке цифровой автоматической телефонной станции (ЦАТС) для выполнения «Требований по технической защите информации…», утвержденных приказом ФСТЭК № 025 от 20.10.2016, вступивших в силу с 01.12.2017. Проанализированы комплектующие из состава ЦАТС для локализации возможных участков с повышенным уровнем электромагнитного излучения. Рассмотрен вариант доработки изделия, позволяющий добиться выполнения требований по защите информации.

Ключевые слова: цифровая автоматическая телефонная станция, электромагнитные излучения, защита информации, канал, сеть передачи данных

PROBLEMS OF SECURITY OF CATS COMMUNICATION CHANNELS IN DATA TRANSMISION NETWORKS

A. V. Serikov[1], L. V. Kuts[2], D. M. Zinoviev[3], R. Yu. Romanikhin[4], S. A. Guzhova[5]

[1,2,3,4,5] *Scientific-industrial Enterprise «Rubin», Penza*

Abstract. The paper deals with the main issues of finalizing the digital automatic telephone exchange (DATS), to fulfill the «Requirements for the technical protection of information …», approved by order of the FSTEC No. 025 dated 10/20/2016, which entered into force on 12/01/2017. Components from the composition of the CATS are analyzed to localize possible areas with an increased level of electromagnetic radiation. A variant of the product refinement is considered, which allows to achieve the requirements for information protection.

Keywords: digital automatic telephone exchange, electromagnetic radiation, data protection, channel, data transmision networks

Введение

Современный этап развития общества характеризуется существенным возрастанием понимания роли и актуальности проблем обеспечения безопасности во всех сферах жизнедеятельности, особенно в сфере информационной безопасности. Повышаются

требования, предъявляемые к оборудованию для защиты информации.

В настоящее время остаются актуальными проблемы, связанные с обеспечением требований по защите информации к мобильным объектам вычислительной техники (ОВТ) – программно-аппаратным комплексам (ПАК) автоматизированного рабочего места (АРМ) в части технических каналов утечки информации за счет побочных электромагнитных излучений (ПЭМИ).

АО «НПП «Рубин» разрабатывает и изготавливает программно-аппаратные комплексы, в состав которых входят ЦАТС [1], в интересах различных структур, в том числе и государственных для организации автоматизированного защищенного обмена речевой информацией по IP-сетям с использованием шифровальной аппаратуры [3].

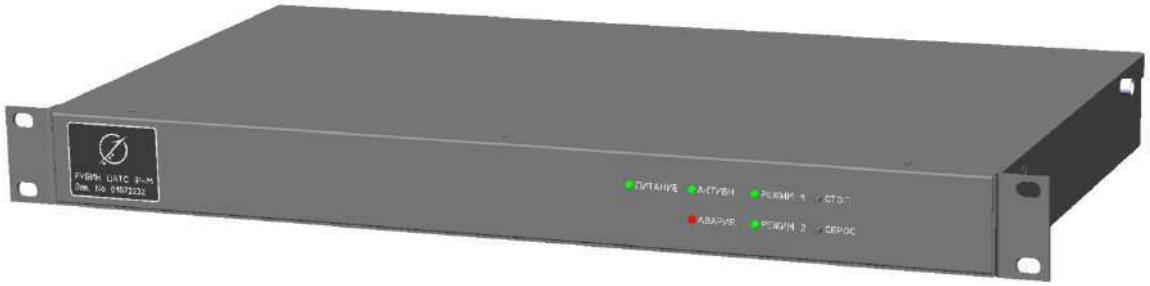
В ходе выполнения работ по доработке изделия ПАК «АРМ «СПЕКТР» [2] в модуле ЦАТС были локализованы участки с повышенным уровнем ЭМИ и проведена работа по их уменьшению, а также решены основные проблемы по обеспечению выполнения требований по защите информации.

В связи с увеличением потребности у Заказчика в количестве абонентских линий до 32, была поставлена задача создания нового изделия на базе ранее разработанного модуля цифровой автоматической телефонной станции (ЦАТС), используемого в ПАК «АРМ «СПЕКТР». Основные технические требования, предъявляемые к разрабатываемому изделию представлены в табл. 1.

Таблица 1

Наименование параметра	Значение
1. Количество портов FXS	До 32 для подключения АТА, в том числе – до 2 факсимильных аппаратов
2. Количество портов Ethernet 10/100 Base-T/TX	2
3. Количество поддерживаемых цифровых ТА	До 32 (с использованием внешнего коммутатора Ethernet)
4. Поддержка кодеков	G.711.a, G.723.1, G.729
5. Протоколы передачи сигнализации и передачи голоса	SIP 2.0, RTP
6. Управление блоком ЦАТС	Через порт Ethernet, посредством программы системы управления (ПО СУ ЦАТС)
7. Мониторинг блока ЦАТС	По протоколу SNMP
8. Наличие системы учёта вызовов	Учёт посредством журнала устанавливаемых соединений
9. Синхронизация даты и времени	1) оператором; 2) посредством протокола NTP

Общий вид блока ЦАТС приведён на рис. 1. Блок реализован в замкнутом металлическом корпусе высотой не более 44 мм (7U) для 19 дюймовой базовой несущей конструкции.



а)

б)

Рис. 1. Общий вид блока ЦАТС:
а – спереди; *б* – сзади

В изделии используется пассивное охлаждение на основе кондуктивного теплоотвода, что позволяет минимизировать количество отверстий и вырезов по периметру корпуса для максимальной эффективности экранирования.

Данная ЦАТС по требованиям заказчика может использоваться как в определенных помещениях капитального строения, так и в транспортных средствах или в различных местах (условиях) эксплуатации. В связи с чем, необходимо выполнить требование к размерам зоны, рассчитанной для портативной возимой аппаратуры разведки ПЭМИ, при использовании ЦАТС в качестве мобильного объекта вычислительной техники.

При проведении измерений ПЭМИ от блока ЦАТС было установлено, что изделие не удовлетворяет требованиям приказа ФСТЭК № 025 к мобильным объектам вычислительной техники.

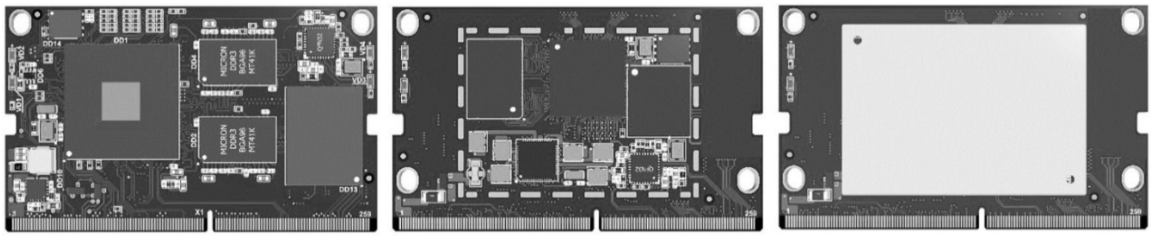
Для локализации «проблемных» участков ЦАТС было принято решение провести спектральный анализ изделия. По результатам спектрального анализа на стенде с помощью сканера Detectus AB RSE644 [4] было выявлено, что источником проблемных

частот в изделии является линия тактирования цифрового звукового интерфейса I2S с несущей частотой 2,048 МГц, гармонические составляющие которой присутствовали в аналоговых телефонных линиях и приводили к недопустимому уровню ЭМИ. Источником тактирования данного интерфейса является система на кристалле (СнК) MCIMX6Q5EYM10AD компании NXP (рис. 2,*а*), а именно внутренний блок генератора частот (PLL). Применение RC фильтров и фильтров серии BLM (Murata) для подавления помех в линиях на печатной плате не дали необходимых результатов, поэтому было принято решение понизить напряжение логического уровня интерфейса I2S с 3,3 В до 1,8 В. Исходя из сформированных требований, а также с целью улучшения производительности системы из-за увеличенного количества аналоговых и цифровых каналов, был разработан АО «НПП «Рубин» новый субмодуль процессора на базе более производительной СнК MIMX8MM6DVTLZAA (рис. 2,*б*) и с возможностью перехода на пониженный логический уровень цифровых интерфейсов. Сравнительные характеристики с ранее разработанным субмодулем, приведены в табл. 2.

Таблица 2

Наименование параметра	Субмодуль на базе MCIMX6Q5EYM10AD	Субмодуль на базе MIMX8MM6DVTLZAA
Тактовая частота, МГц	800	1800
Тип и объем оперативной памяти	DDR3L, 1 Гбайт	LPDDR4, 2 Гбайта
Разрядность шины адреса	32	64
Габариты корпуса СнК, мм	21×21	14×14
Потребляемая мощность	4 Вт	2 Вт

Новая СнК имеет более производительный интерфейс оперативной памяти LPDDR4 с пониженным напряжением питания в 1,1 В, что положительно сказывается на уровне излучаемых помех. Меньшие габариты корпуса и использование компонентов типоразмером 0201 позволили уменьшить площадь трассировки печатной платы и поместить все элементы в экранирующий кожух SMS-210F (Leader Tech). Данный кожух (рис. 2,*в*) совместно с опорными слоями многослойной печатной платы обеспечивают эффективное экранирование со всех сторон, в том числе уменьшает ЭМИ генератора частот СнК.



а) *б*) *в*)

Рис. 2. Внешний вид субмодулей процессора:
а – на базе MCIMX6Q5EYM10AD; *б* – на базе MIMX8MM6DVTLZAA;
в – на базе MIMX8MM6DVTLZAA с экранированным кожухом

Переход на логический уровень в 1,8 В интерфейса I2S позволил существенно снизить уровень ЭМИ, однако цифровые линии стали менее устойчивыми к внутренним источникам помех, таким как источники питания и преобразователи напряжения. Это привело к частичным потерям данных, искажению голоса, что негативно сказывалось на стабильности работы изделия и качестве связи.

В ходе электромагнитного сканирования были выявлены наиболее выраженные источники помех (рис. 3), которыми оказались полевые транзисторы, применяемые в качестве ключа в инверсных повышающих преобразователях. Данные преобразователи используются для независимого питания каждого телефонного канала, формирования тонового сигнала и звонкового напряжения.

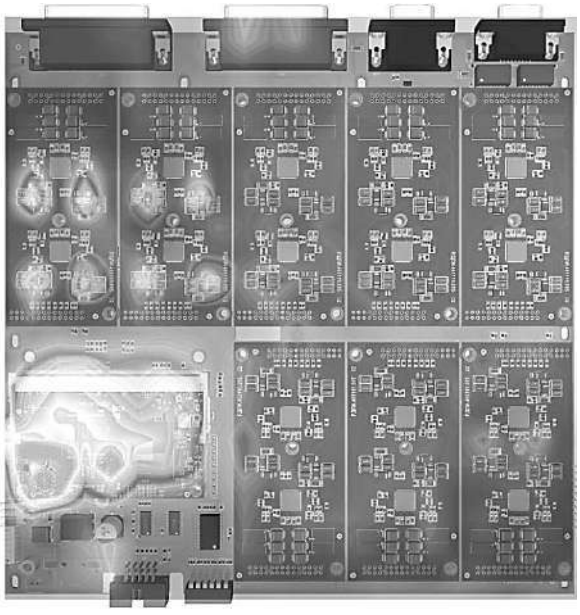


Рис. 3. Спектральный анализ модуля ЦАТС

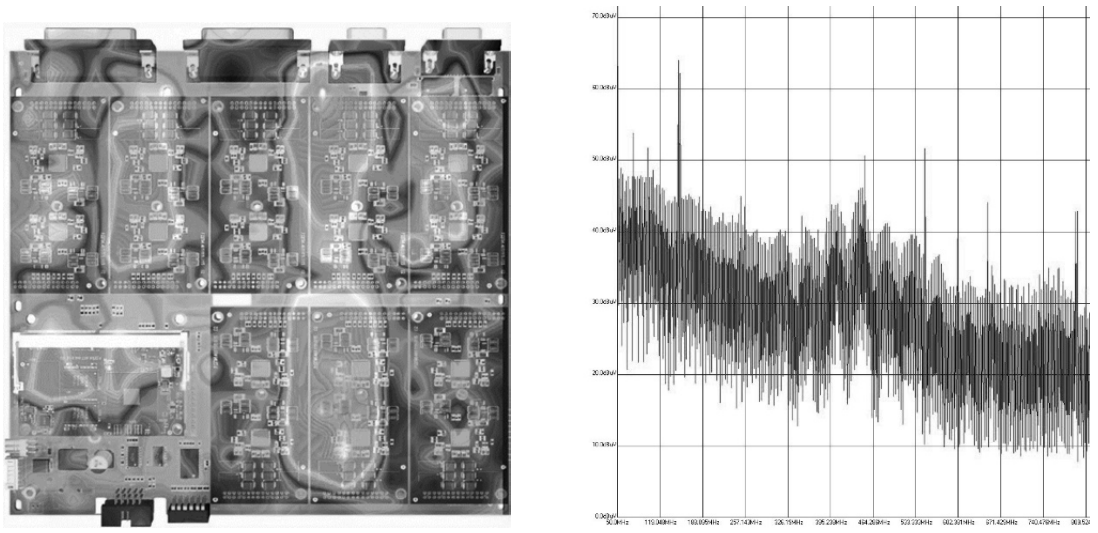
Для уменьшения негативного воздействия на линии I2S была проведена доработка инверсного повышающего преобразователя, где основным источником помех являлся полевой транзистор

STD5N20L (ST). В связи с конструктивными особенностями данного типа транзисторов, затвор обладает паразитными параметрами емкости и индуктивности, что приводит к высокочастотному звону, который передается в цепи питания. Для демпфирования данного звона в цепь затвора помимо токоограничительного резистора был добавлен BLM фильтр, а также осуществлена замена самого полевого транзистора STD5N20L (ST) на FDT86246L (OnSemi) с более лучшими характеристиками. Помимо меньших паразитных параметров, новый транзистор в открытом состоянии обладает малым сопротивлением (228 мОм) между стоком и истоком, что в свою очередь уменьшило потребляемую мощность и положительно сказалось на температурных характеристиках изделия.

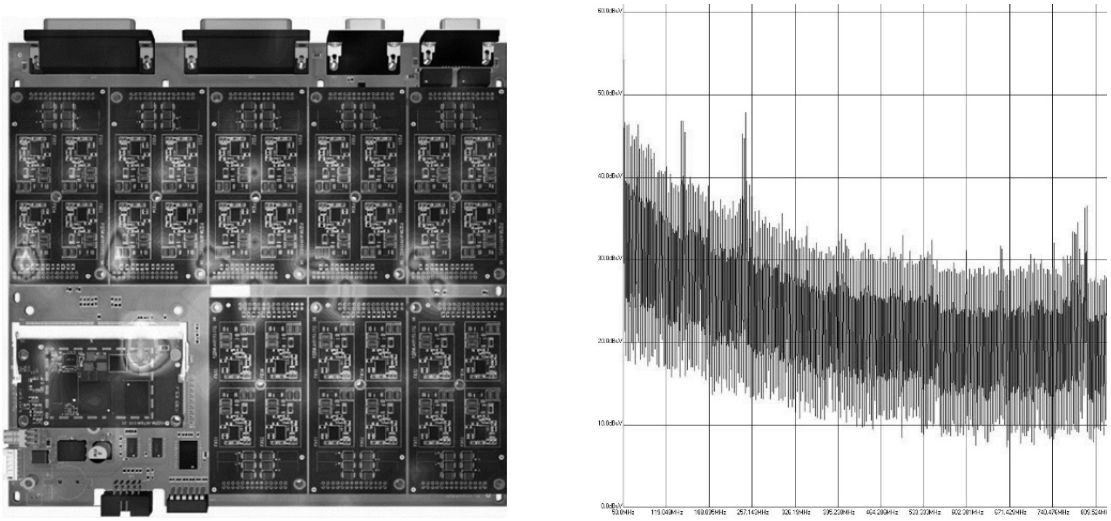
Управление транзисторным ключом повышающего преобразователя осуществляется с помощью двухканальной микросхемы телефонного интерфейса (SLIC) Le9632 которая преобразует цифровой сигнал интерфейса I2S в аналоговый сигнал телефонной линии. Данные микросхемы применялись в ранее разработанном модуле ЦАТС из-за упрощенного управления по линии SPI по сравнению с одноканальными. Однако, сравнительный спектральный анализ и функциональное тестирование показали, что аналогичные одноканальные микросхемы Le9653 оказывают меньше негативного воздействия на интерфейс I2S (рис. 4). Это обусловлено тем, что микросхема Le9632 имеет совмещенный блок управления двумя источниками питания телефонной линии, и из-за близкого расположения источников ШИМ сигнала и цифрового звукового интерфейса. Это неизбежно приводит к большим помехам, чем с использованием одноканальных микросхем. В связи с этим, в разрабатываемой ЦАТС была также осуществлена доработка, связанная с заменой микросхем телефонного интерфейса на Le9653.

По завершению работ был повторно проведен спектральный анализ и проведены измерения уровня ПЭМИ. Результаты подтвердили эффективность применения указанных выше доработок.

Таким образом, разработка нового субмодуля процессора, переход на пониженный логический уровень звукового интерфейса, доработка внутренних источников питания и переход к одноканальным микросхемам телефонного интерфейса позволили выполнить требования по технической защите информации к мобильным объектам вычислительной техники с учетом увеличения числа абонентских линий, а следовательно, были улучшены характеристики изделия.



а)



б)

Рис. 4. Сравнительный анализ на стенде ЭМС:
а – результаты сканирования до доработки; *б* – после доработки

Данный комплекс мероприятий позволил обеспечить низкий уровень ПЭМИ внутренних блоков изделия, что позволило использовать простой и недорогой корпус для ЦАТС без применения дорогостоящих экранирующих материалов.

Список литературы

1. Акимов М. В., Бабич А. М. Переносные программно-аппаратные комплексы автоматизации и связи // Проблемы информатики в образовании, управлении, экономике и технике : сб. ст. XVIII Междунар. науч.-техн. конф., посвящ. 75-летию Пензенского государственного университета. Пенза : Изд-во ПГУ, 2018. С. 128–133.

2. Зиновьев Д. М., Куц Л. В., Пашкин А. В. [и др.]. Специальная тема // Информационно-управляющие, телекоммуникационные системы,

средства поражения и их техническое обеспечение : сб. науч. ст. по материалам IV Всерос. межведомственной науч.-техн. конф. / под общ. ред. М. М. Бутаева. Пенза : АО «НПП "Рубин"», 2022. С. 80–86.

3. Куц Л. В., Шариков А. В., Говор Т. А. [и др.]. Разработка и интеграция программно-аппаратного комплекса автоматизированной коммутации абонентов телефонной сети специальной связи // Информационно-управляющие, телекоммуникационные систем, средства поражения и их техническое обеспечение : сб. науч. ст. по материалам III Всерос. межведомственной науч.-техн. конф. Пенза, 2021. С. 138–148.

4. Востоков Н. Г., Горбунов А. С. Снижение электромагнитного излучения электронного устройства с помощью технологии размытия спектра // Известия высших учебных заведений. Поволжский регион. Технические науки. 2020. № 2 (54). С. 54–64.

УДК 004.056

ПОВЫШЕНИЕ БЕЗОПАСНОСТИ ИНФОРМАЦИИ В КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЯХ

И. В. Чёрный

Научно-производственное предприятие «Рубин», г. Пенза

Аннотация. Рассматривается актуальность проблемы безопасности информационных технологий на современном этапе. Доводятся до сведения читателей положения федерального закона «Об информации, информационных технологиях и о защите информации». Представлены информационные технологии как таковые и их безопасность с позиции применения в профессиональной и повседневной деятельности. Приводятся факторы определяющие, применяющиеся средства, а также мероприятия, проводящиеся для обеспечения сохранности и защиты информации. Предлагается на рассмотрение направление повышения информационной безопасности технологий, основанное на представлении и хранении информации в формульном виде с описанием её в формате естественного языка.

Ключевые слова: современные информационные технологии, безопасность информационных технологий, защита информации, формульное представление информации, естественный язык, естественное представление информации

IMPROVING INFORMATION SECURITY IN COMPUTER TECHNOLOGY

I. V. Cherny

Scientific-industrial Enterprise «Rubin», Penza

Abstract. The article discusses the relevance of the problem of information technology security at the present stage. The provisions of the federal law «On Information, information Technologies and Information Protection» are brought to the attention of readers. Information technologies as such and their security from the point of view of application in professional and daily activities are presented. The determining factors, the means used, as well as the measures taken to ensure the safety and protection of information are given. The direction of improving the information security of technologies based on the presentation and storage of information in a formulaic form with its description in natural language format is proposed for consideration.

Keywords: modern information technologies, information technology security, information protection, formulaic representation of information, natural language, natural representation of information

Стремительное развитие и распространение современных информационных технологий, в настоящее время приобретает характер глобальной информационной революции, оказывает все более возрастающее влияние практически на все сферы жизнедеятельности общества и привели к активному обсуждению теоретических аспектов развития безопасности информационных технологий.

Федеральный Закон РФ от 27 июля 2006 г. № 149-ФЗ «Об информации, информационных технологиях и о защите информации» определяет, что целями защиты информации являются: предотвращение утечки, хищения, утраты, искажения, подделки информации; предотвращение несанкционированных действий по уничтожению, модификации, искажению, копированию, блокировке информации [1].

Защищённость информационных технологий от случайных или преднамеренных воздействий естественного или искусственного характера, способных нанести ущерб владельцам и пользователям информации её структуре и составу, состоит из основных элементов: доступность-обеспечение готовности системы к обслуживанию поступающих к ней запросов; целостность-представление настоящей информации в неизменном виде; конфиденциальность-защита информации от несанкционированного проникновения к её ресурсам [2].

Информационные технологии упрощают нашу жизнь, помогают в проделывании коллективной работы, способствуют быстрой коммуникации на расстоянии, помогая сохранить личное время. Проблема защиты информационных технологий в современных условиях актуальна и обуславливается основными факторами такими как: расширение сферы использования ЭВМ, распространением разнообразных информационно-управляющих систем, повышения уровня доверия к ним и использованием их в опасных областях деятельности, привлечением в процесс информационного взаимодействия большого количества людей и организаций, возрастание их информационных потребностей, объединение огромных массивов информации различного назначения, усовершенствование способов доступа к информационным ресурсам, возникновение и обострение противоречий между потребностями общества, развитием свободного обмена информацией и недостаточной защищенностью её использования и распространения, расширения угроз незаконного доступа к информации, прирост компетентных пользователей вычислительных машин способных

к созданию отрицательных программно-математических воздействий на системы обработки информации, возрастание потерь от подделывания, уничтожения, искажения, разглашения и несанкционированного распространения информации [2].

Сегодня всемирная сеть интернет является самый востребованный ресурс по поиску какой-либо информации и её интерпретированию при этом она же представляет большую угрозу и является необозримо опасным каналом для каждой машины. Специалисты стремятся как можно больше приложить усилий в области создания и разработки различных программных обеспечений, антивирусных систем, криптограмм и многого другого для защиты уязвимой информации. Для недопущения «нештатных» ситуаций, вниманию человека предлагаются различные средства и организационные мероприятия для защиты. Использование различных средств защиты информации в виде инженерно-технических, электрических, электронных, оптических устройств и приспособлений, помогает в решении различных задач по защите информации и предупреждения её утечки. При этом во всех при рассмотрении проблемы информационной безопасности, используются понятия «несанкционированный доступ» и «санкционированный доступ». Где первое предполагает неправомочное обращение к информационным ресурсам с целью их использования, присвоения, порчи или уничтожения. А второе предполагает обращение к объектам, программам и данным, выполнение действий с информацией (чтение, обработку и т.д.), а также использование ресурсов и услуг, при наличии разрешения администратора вычислительной системы [3].

В зависимости от способа реализации, средства обеспечения защиты информации подразделяют:

– технические (аппаратные) средства, при помощи аппаратных систем определяют задачи по защите информации за счет блокирования входа физическим лицам. Надежность, независимость от субъективных факторов, высокая устойчивость к модификации являются преимуществом таких средств, а недостаточная гибкость, относительно большие объем и масса с высокой стоимостью являются слабой стороной;

– программные средства выполняющие функции по распознаванию и идентификации пользователей, контролю доступа, шифрования, уничтожению непригодной информации. Их преимущество заключается в универсальности, гибкости, надежности, простоте установки, способности к модификации и развитию,

а ограниченная функциональность сети, использование части ресурсов файл-сервера и рабочих станций, высокая чувствительность к случайным или преднамеренным изменениям, возможная зависимость от типов компьютеров являются их недостатками;

– организационные состоят из организационно-технических и организационно-правовых средств. Их плюс в том, что они позволяют решать множество разнородных проблем, они примитивны в реализации, имеют неограниченные возможности модификации и развития, при этом в большинстве своем зависят от субъективных факторов;

– криптографические средства используют шифрование, кодирование информации, приводя к тому, что информация или ресурсы перестают быть доступны без введения специального ключа. Это средство применяется как программа и считается самым надежным, так как непосредственно защищена информация, а не доступ к ней [4].

Проводимые для обеспечения сохранности и защиты информации организационные мероприятия предполагают, объединение всех составляющих безопасности информации, подразделяемых:

– на социально-психологические предполагающих проведение регулярного мониторинга организационных мероприятий по недопущению отрицательных воздействий и явлений;

– физические подразумевающие использование человеческих ресурсов, специально оборудованных технических средств и устройств, позволяющих гарантировать защиту информации от проникновения злоумышленников на объект, преступного использования, порчи или уничтожения ими материальных и нематериальных ресурсов.

Злоумышленники используют характерную особенность электронных данных, связанную с возможностью их непринуждённого и незаметного копирования, искажения или уничтожения их из базы данных, в которой они хранятся. В связи с этим существует постоянная необходимость непрерывного совершенствования организации безопасного функционирования данных в любых информационных системах, стараясь максимально разработать системы и средства для борьбы со злоумышленниками и угрозами, которые могут нанести ущерб.

Все вышесказанное позволяет свести к минимуму намерения, направленные на взлом информации с целью воспользоваться ею в своих интересах или нанести ущерб информационному каналу. Но при этом все равно остаются возможности к достижению

этих целей так как даже на этапе разработки различных программных сервисов или программно-технических устройств непреднамеренно или намеренно (для создания задела на новые разработки) закладываются предпосылки к этому. На мой взгляд разработчики средств защиты информации всегда будут в роли догоняющих по отношению к взломщикам, а целостность и конфиденциальность информации все равно будет оставаться под угрозой. Дело в том, что информация в ЭВМ по своей сути является открытой. Хранение её также осуществляется по большому счёту в открытом виде. А все средства, обеспечивающие её защиту (технические, программные, организационные и криптографические) это всего лишь преграды, которые необходимо преодолеть злоумышленнику что бы украсть, изменить или уничтожить информацию.

Таким образом можно сделать вывод что бы обеспечить необходимый уровень безопасности информационных технологий в современных информационных системах необходимо начать с основ представления, организации и хранения информации и способов её обработки. Чтобы при случайных или преднамеренных воздействиях естественного или искусственного характера была бы полностью исключена возможность нанесения ущерба владельцам и пользователям информации её структуре и составу, а также её использованию.

При этом должна обеспечиваться в информационной системе способность владельца информации и ресурсов контролировать (разрешать, запрещать) доступ к ней потребителей. Наиболее эффективно это достигается если информация перед её представлением не будет где-то браться (например, из базы данных, каталога и т.д.) для предоставления, а формироваться по определенному алгоритму.

Так как информация является не только сведениями об объектах и явлениях окружающей среды, их параметрах, свойствах и состоянии, которые уменьшают имеющуюся о них степень неопределенности, неполноты, но и та часть знаний, которая используется для ориентирования, активного действия, управления, в целях сохранения и развития системы. То для хранения информации из различных областей знаний необходима единая технология (формат) её представления.

В качестве такой технологии необходимо использовать формулы как вид информационного моделирования. За счет того, что они встречаются во многих областях знаний математике, физике,

химии, экономике, статистике, логике и т.д. они являются универсальным механизмом для описания информации различных областей знаний. Основное преимущество формул от других способов представления информации состоит в том, что информация показана в них в наиболее «свернутом», компактном виде и позволяет:

– дать описание объекта в наиболее компактном виде;

– отразить причинно-следственные связи физического явления;

– передать такие свойства объекта, которые не поддаются описанию другими средствами;

– предсказать свойства и поведение моделируемого объекта за пределами видимых наблюдений [5].

В формулах практически нет избыточной информации, не только каждый знак, но и их взаимное расположение несут важную смысловую нагрузку. При использовании в качестве «знаков» формул «символов» естественного языка можно описать любую информацию от простого сообщения до объяснения устройства какой-нибудь технической системы. Естественный язык достаточно гибкий:

– естественный язык имеет очень широкую сферу применения;

– при помощи естественного языка можно передавать любую информацию;

– в естественном языке есть большое количество правил;

– естественный язык имеет явный характер, например правила грамматики;

– естественный язык достаточно открыты, в нем можно образовывать новые слова, то есть в нем развиты механизмы словообразования;

– естественный язык способен динамично развиваться и подстраиваться под потребности взаимодействия, это видно на примере различных профессиональных диалектов [6].

Совокупность правильно построенных формул какой-либо формальной системы позволяет создать логические модели знаний, обеспечивающих:

– высокий уровень формализации, обеспечивающий возможность реализации системы формально точных определений и выводов;

– согласованность знаний как единого целого, облегчающая решение проблем верификации базы знаний;

– предоставление единого средства описания как знаний для различных предметных областей, так и способов решения задач

в этих предметных областях, позволяет любую задачу свести к построению и последующему логическому анализу формул одной или последовательностей взаимосвязанных формул из нескольких формальных систем логических моделей знаний [5].

Для преобразования информации в формулы с описанием в них её на естественном языке отсутствует необходимость разработки множества алгоритмов её анализа и преобразования, а достаточно всего одного независимо на каком языке (русский, английский и т.д.) она представлена в виду того что в данном случае нам её необходимо сохранить обеспечив целостность и безопасность.

Такое представление информации позволяет воспользоваться ею только при наличии специального программного обеспечения способного прочитать и преобразовать информацию, представленную в формульном виде для последующего её использования различными программными сервисами. Формульное представление информации сродни криптографическим средствам с разностью в том, что вместо ключа используется специальное программное обеспечение совместно с информацией о допуске к её использованию находящейся в самой информации. Это позволяет даже в случае хищения или попытке порчи или уничтожения информации в том числе в случае наличия специального программного обеспечения сохранить целостность и конфиденциальность информации.

Данный подход не ограничивается только обеспечением безопасности информации. Формульное её представление обеспечивает полную формализацию данных, а естественное представление способствует облегчению её передачи (разработка единого протокола) и созданию наукоемкого программного обеспечения, осуществляющего взаимную обработку информации из различных областей знаний. Что в свою очередь расширяет возможности по созданию информационных, управляющих и других программных средств и систем в стремительном развитии и распространении современных компьютерных технологий как для продукции выпускаемой непосредственно АО «НПП «Рубин», так и для других сфер жизнедеятельности общества.

Список литературы

1. Об информации, информационных технологиях и о защите информации : федер. закон РФ № 149-ФЗ от 27 июля 2006 г. URL: https://cga.mos.ru/upload/medialibrary (дата обращения: 01.09.2023).

2. Тихонова А. В. Информационные технологии и безопасность // Перспективы развития информационных технологий. Компьютерные и информационные науки. 2013. № 12. С. 228–233.

3. Гайсинский И. Е., Вострикова Т. В., Перова М. В., Зверяко А. Е. Информационные технологии : учеб. пособие / под общ. ред. И. Е. Гайсинского. Ростов н/Д., 2012.

4. Галатенко В. А. Основы информационной безопасности : учеб. пособие / под ред. акад. РАН В. Б. Бетелина. 4-е изд. М. : Интернет-Университет информационных технологий : Бином. Лаборатория знаний, 2008.

5. Глотова М. Ю. Математическое моделирование как один из основных методов познания // Математическая обработка информации. URL: https://studme.org/281961/matematika_himiya_fizik/matematicheskoe_modelirovanie_osnovnyh_metodov_poznaniya (дата обращения: 01.09.2023).

6. Снежкова Л. Язык и алфавит представления информации // Образовательный портал «Справочник». URL: https://spravochnick.ru/informatika/kodirovanie_informacii/yazyk_i_alfavit_predstavleniya_informacii/ (дата обращения: 02.09.2023).

УДК 004.032.26

ФОРМИРОВАНИЕ ОБУЧАЮЩЕЙ ВЫБОРКИ НА ОСНОВЕ ТЕКСТОВЫХ ДАННЫХ ДЛЯ ПСИХОЛОГИЧЕСКОГО АНАЛИЗА ПОЛЬЗОВАТЕЛЕЙ С ИСПОЛЬЗОВАНИЕМ НЕЙРОННЫХ СЕТЕЙ

А. М. Бабич[1], Ю. В. Шарикова[2], А. А. Зоткина[3], Е. Н. Семёнова[4], М. Ю. Уткина[5]

[1,2,4,5] Научно-производственное предприятие «Рубин», г. Пенза
[3] Пензенский государственный технологический университет, г. Пенза

Аннотация. На сегодняшний день основополагающими способами определения отклоняющего поведения являются методы, основанные либо на психологических тестах, либо на непосредственной формализации врача. Такой способ сложен и относительно дорог, кроме того, он требует согласия со стороны пациента. Решением данной проблемы является использование методов машинного обучения, в частности, активно развивающиеся в настоящее время методы использования нейронных сетей. Целью данного исследования является подготовка обучающей выборки на основе больших объёмов исходных текстовых и скалярных данных для последующего обучения нейронных сетей, предназначенных для анализа пользователей на наличие отклоняющего поведения. Для решения данной задачи удобно использовать социальные сети ввиду наличия сообществ, основной тематикой которых является обсуждение проблем, связанных с психологическим состоянием пользователей. В ходе работы был выбран тип архитектуры нейронной сети и сформирована база знаний для обучения.

Ключевые слова: социальная сеть, анализ публикаций, текстовая информация, большие данные, обучающая выборка, нормализация данных, отклоняющее поведение, нейронная сеть

CREATING OF A TRAINING SAMPLE BASED ON TEXT DATA FOR USER PSYCHOLOGICAL ANALYSIS USING NEURAL NETWORKS

A. M. Babich[1], Yu. V. Sharikova[2], A. A. Zotkina[3], E. N. Semyonova[4], M. Yu. Utkina[5]

[1,2,4,5] Scientific-industrial Enterprise «Rubin», Penza
[3] Penza State Technological University, Penza

Abstract. Today, the fundamental methods for determining deviant behavior are methods based either on psychological tests or on the direct formalization of

a doctor. This method is complex and relatively expensive; in addition, it requires the consent of the patient. The solution to this problem is the use of machine learning methods, in particular, the currently actively developing methods of using neural networks. The purpose of this study is to prepare a training sample based on large volumes of source text and scalar data for subsequent training of neural networks designed to analyze users for the presence of deviant behavior. To solve this problem, it is convenient to use social networks due to the presence of communities, the basis of which is the discussion of problems related to the psychological state of users. During the work, the type of neural network architecture was selected and a knowledge base for training was created.

Keywords: social network, publication analysis, text information, big data, training set, data normalization, deviating behavior, neural network

Основное предназначение социальных сетей – предоставление средств для общения, знакомства и обмену открытой информацией между пользователями. К такой информации можно отнести сведения, доступные в текстовых публикациях (а также сама частота публикаций), сведения о личных интересах, содержание фотографий. Таким образом, социальные сети являются удобным интернет-ресурсом, обеспечивающим накопление данных, отражающих особенности личности пользователя и позволяющих создать его психологический портрет. Кроме того, социальные сети объединяют людей со схожими интересами, что позволяет выделять группы людей по некоторому социальному или личностному признаку. Учитывая, что социальные сети вовлекают в себя десятки и сотни миллионов пользователей, они являются одним из основных ресурсов, обеспечивающих сбор и накопление больших данных, пригодных для создания обучающих выборок для искусственных нейронных сетей.

На сегодняшний день поведенческие и лингвистические маркеры, извлекаемые из социальных сетей, применяются в исследованиях, посвященных определению настроения и психологических расстройств пользователей. С 2013 года были проведены исследования, на основе которых сформированы методы, позволяющие с высокой точностью определить развитие шизофрении, тяжелой депрессии, склонности к суициду, и расстройств пищевого поведения [1]. Данные маркеры были получены посредством анализа истории публикаций и поведенческих особенностей пользователей в таких социальных сетях, как Twitter, Reddit и Facebook (признана экстремистской организацией и запрещена на территории РФ).

Извлекаемая подобным образом информация может использоваться для формирования обучающих выборок при решении таких задач как дополнение и уточнение данных клинического обследования, улучшение результатов психиатрической помощи, сбора статистических данных для исследования настроений в обществе, выявления опасного поведения, в том числе в случаях, когда обычные клинические подходы неприменимы. Также согласно [2] подобные подходы использовались платформой Facebook (признана экстремистской организацией и запрещена на территории РФ) для предотвращения самоубийств.

Методы проведения таких исследований берутся из междисциплинарных областей, таких как машинное обучение, медицинская информатика, искусственный интеллект, обработка текстов на естественном языке и взаимодействие человека и компьютера.

Данная работа посвящена подготовке инструментария, сбору и предварительной обработке информации для формирования обучающей выборки, используемой при последующем психологическом анализе пользователей социальных сетей на наличие отклоняющего поведения. В качестве базового интернет-ресурса выбрана социальная сеть ВКонтакте. Данная задача схожа с приведёнными ранее задачами определения настроения и психологических расстройств пользователя и решается аналогичными методами.

Согласно [3] в качестве средства анализа пользователей социальных сетей наиболее эффективно использовать искусственные нейронные сети (ИНС), специализирующиеся на решении задач классификации и обобщения информации. В роли признаков для обучения ИНС, как правило, используются две группы исходных данных:

– численные статистические параметры со страницы пользователя (скаляры), например число подписчиков, число друзей, число подарков;

– текстовые параметры – последовательность слов в публикации и в статусе. При этом эмодзи-символы заменяются соответствующим текстовым описанием.

Для сбора информации из социальной сети ВКонтакте использовался официальный API приложений. Был произведён анализ методов данного API для извлечения данных из профилей пользователя или группы. В табл. 1 приведена общая информация об анализируемых параметрах и соответствующие методы API ВКонтакте, осуществляющие доступ к базе данных.

Таблица 1

Сведения о параметрах для обучения ИНС

Параметр	Метод	Тип исходных данных
1. Список записей со стены пользователя или сообщества	wall.get	Последовательность слов
2. Текст статуса пользователя или сообщества	status.get	Последовательность слов
3. Расширенная информация о пользователях (в том числе число подписчиков)	users.get (параметр followers_count)	Скаляр
4. Список идентификаторов друзей пользователя или расширенная информация о друзьях пользователя (в том числе общее число друзей)	friends.get (параметр count)	Скаляр
5. Список полученных подарков пользователя (в том числе общее число подарков)	gifts.get (параметр count)	Скаляр

Для осуществления автоматизированного сбора исходных данных был получен токен ВКонтакте (ключ доступа). При помощи токена приложение сообщает серверу, от имени какого пользователя осуществляются запросы, и какие права доступа ему предоставлены. Официальный API ВКонтакте предоставляет несколько типов токенов. В данной работе был использован тип «ключ доступа пользователя». Данный ключ используется для работы со всеми открытыми методами API ВКонтакте (кроме методов секции secure) [4]. Этот набор методов полностью удовлетворяет требованиям для автоматизированного сбора данных.

Для сбора информации и подготовки обучающей выборки использовался скриптовый язык программирования Python. Такой выбор обусловлен тем, что он широко применяется в задачах, связанных с большими данными и созданием ИНС, и соответственно, обладает широким набором соответствующих инструментов [5].

Для взаимодействия с API ВКонтакте использовалась библиотека requests, являющаяся стандартным инструментом для составления HTTP-запросов в Python [6].

На этапе предварительной обработки текста и нормализации данных использовалась специализированная библиотека NLTK (Natural Language Toolkit). Это ведущая платформа для решения задач по обработке естественного языка (NLP-программы) на языке Python. Эта библиотека предоставляет инструментарий для обработки текста на большом количестве языков и решения таких задач по обработке естественного языка как токенизация, разметка,

кластеризация, фильтрация и т.д. На этом этапе текстовые данные очищаются от знаков препинания, ссылок и специальных символов. Числа и эмодзи-символы заменяются их словесным описанием.

Как указывалось выше, в качестве средства психологического анализа пользователей социальных сетей были выбраны ИНС. ИНС – это математическая модель, имитирующая работу клеток мозга и имеющая, как правило, программную реализацию. Способность ИНС к обучению дает преимущество перед традиционными алгоритмами распознавания. В процессе обучения сеть выявляет зависимости между данными, а также способна выполнять обобщение. ИНС способна выдавать верные результаты на основе данных, не присутствующих в обучающей базе.

Вид ИНС определяется задачами, которые данная сеть должна решать в ходе своей работы. Поскольку в данном случае базовой анализируемой информацией являются текстовые данные, длина которых варьируется, то наиболее подходящим видом нейронной сети являются рекуррентная нейронная сеть (РНС), поскольку такой класс нейронных сетей используется в задачах обработки последовательностей нефиксированной длины [7]. В частности, рекуррентные нейронные сети показывают результаты лучше других методов в задачах классификации текста [8]. Рекуррентные нейронные сети – это подкласс нейронных сетей с обратными связями, которые используют предыдущие состояния сети для вычисления текущего. Реализация РНС в данной работе осуществляется при помощи библиотеки Keras. Структура и размер РНС определяются на основе объёма данных в обучающей выборке – чем больше данных, тем больше весовых коэффициентов можно определить, а также исходя из сложности задачи (определяется экспериментально).

В качестве исходных групп для обучения ИНС был подобран следующий список:

– psyhelp_online («Help | Психологическая помощь»);

– psymedcareru («Панические атаки ВСД ПА фобии неврозы общение»);

– muesli_na_ventilatore («Мюсли на вентиляторе»);

– gosha_froll («Исповедь суицидника 18+»);

– caps__rage.

В данный список были внесены группы, обладающие значимым количеством текстовых сообщений, характерных для психологического состояния пользователей с отклоняющим поведением.

В формируемую выборку для последующего обучения ИНС добавлялась как нормализованная текстовая информация, так и доступная статистическая информация из профилей подписчиков и комментаторов группы. Был создан скрипт Python по автоматизированному сбору текстовых и числовых статистических данных из указанных групп. Данные были нормализованы и сохранены для последующего обучения РНС.

Таким образом, в данной работе был определён метод психологического анализа аудитории социальной сети ВКонтакте, определены инструменты для сбора и очистки исходных данных для формирования обучающей выборки, выбрана библиотека для обеспечения взаимодействия с официальным API социальной сети. Определён тип и произведён сбор исходных данных для обучения ИНС. Для обучения выбран рекуррентный тип ИНС, поскольку данные нейронные сети применимы для обработки текстовой информации переменной длинны и активно используются для схожих задач по анализу текстов с психологической точки зрения [9].

В качестве дальнейшего направления работ можно выделить расширение исходной базы данных путём поиска новых групп соответствующей тематики, непосредственное обучение РНС после сбора достаточного количества материала для глубокого обучения и коррекция обучающей выборки по результатам обучения.

Список литературы

1. Chancellor S., De Choudhury M. Methods in predictive techniques for mental health status on social media: a critical review // NPJ Digital Medicine. 2020. Vol. 3 (1). P. 1–11.

2. Vincent J. Facebook is using AI to spot users with suicidal thoughts and send them help. URL: https://www.theverge.com/2017/11/28/16709224/facebook-suicidalthoughts-ai-help (дата обращения: 31.08.2023).

3. Браницкий А. А., Дойникова Е. В., Котенко И. В. Использование нейросетей для прогнозирования подверженности пользователей социальных сетей деструктивным воздействиям // Информационно-управляющие системы. 2020. № 1. С. 24–33.

4. VK для разработчиков. Общие сведения. Ключ доступа пользователя. URL: https://dev.vk.com/api/access-token/getting-started (дата обращения: 31.08.2023).

5. Шолле Ф. Глубокое обучение на Python. СПб. : Питер, 2018. 400 с.

6. Requests в Python – примеры выполнения HTTP запросов. URL: https://python-scripts.com/requests?ysclid=lgxqbf6l5r84638924#method-get-requests (дата обращения: 31.08.2023).

7. Зоткина А. А. Рекуррентные нейронные сети как алгоритм последовательности данных // Современные информационные технологии. 2022. № 35. С. 24–26.

8. Lai S., Xu L., Liu K., Zhao J. Recurrent Convolutional Neural Networks for Text Classification // AAAI. 2015. P. 2267–2273.

9. Alireza Souri1, Shafgheh Hosseinpour and Amir Masoud Rahmani. Personality classifcation based on profles of social networks' users and the five-factor model of personality // Souri et al. Hum. Cent. Comput. Inf. Sci. 2018. URL: https://doi.org/10.1186/s13673-018-0147-4 (дата обращения: 31.08.2023).

УДК 004.056

ИСПОЛЬЗОВАНИЕ НАБОРА НЕЙРОСЕТЕВЫХ ВАРИАНТОВ СТАТИСТИЧЕСКИХ КРИТЕРИЕВ ПРИ ОЦЕНКЕ МАЛЫХ БИОМЕТРИЧЕСКИХ ВЫБОРОК

С. А. Гужова[1], А. И. Иванов[2], Н. А. Папуша[3], А. С. Кирин[4], М. А. Ерков[5]

[1] Научно-производственное предприятие «Рубин», г. Пенза
[2] Пензенский научно-исследовательский электротехнический институт, г. Пенза
[3,4,5] Пензенский государственный университет, г. Пенза

Аннотация. Рассмотрена одна из проблем в обучении нейронной сети, а точнее – необходимость для этого большой обучающей выборки. Приведен обзор статистического критерия Андерсона – Дарлинга и Пирсона, представлены их эквивалентные формы в виде искусственных нейронов. Показано уменьшение вероятностей ошибок первого и второго рода путем дифференцирования данных критериев. Предложен переход от нейросетевого анализа данных в статике к анализу в динамике для увеличения правильности принимаемых нейросетью решений.

Ключевые слова: нейронная сеть, критерий хи-квадрат, критерий Андерсона – Дарлинга, малые выборки, нейродинамика

USING A SET OF NEURAL NETWORK VARIANTS OF STATISTICAL CRITERIA IN THE EVALUATION OF SMALL BIOMETRIC SAMPLES

S. A. Guzhova[1], A. I. Ivanov[2], N. A. Papusha[3], A. S. Kirin[4], M. A. Erkov[5]

[1] Scientific-industrial Enterprise «Rubin», Penza
[2] Penza Scientific Research Electrotechnical Institute, Penza
[3,4,5] Penza State University, Penza

Abstract. One of the problems in training a neural network was considered, or rather the need for a large training sample for this. The statistical criterion of Anderson – Darling and Pearson is reviewed, their equivalent forms in the form of artificial neurons are presented. The reduction of the probabilities of errors of the first and second kind by differentiating these criteria is considered. A transition from neural network data analysis in statics to dynamic analysis was proposed to increase the correctness of decisions made by the neural network.

Keywords: neural network, chi-square criterion, Anderson – Darling criterion, small samples, neurodynamics

Интерес к искусственному интеллекту возник у человечества еще в прошлом веке. Впервые модель искусственного нейронная была предложена в 1943 году, а первая его программная реализация в 1957 году. Спустя практически около 80 лет нейронные сети масштабно внедрились в жизнь человека и обширно применяются практически во всех областях его жизнедеятельности. Они используются для обработки, структурирования и анализа больших объёмов информации в естественных, технических, общественных, гуманитарных науках. Нейронные сети являются мощной вычислительной средой, что безусловно сильно облегчает жизни людей, сокращая время и силы, затраченные бы человеком на эту работу. Одной из проблем использования данного метода вычислений и обработки является обучение этих сетей, а точнее необходимость для этого большой обучающей выборки [1]. На практике не всегда удается или крайне сложно собрать необходимое количество образов, при условии, что стандартные статистические рекомендации прописывают необходимость привлекать свыше 200 образов в обучающей выборке для проверки закона нормального распределения. Например, для биометрической аутентификации, работающей по рукописному слову-паролю или какому-либо другому критерию, человеку приходится ввести необходимый параметр, для обучения нейронной сети, порядка 200 раз. Это неприемлемо много, из-за этого у пользователя складывается отрицательное мнение об этом виде аутентификации и он, как правило, отказывается в его использовании по причине данных сложностей. Однако ситуация в корни изменилась если бы человеку предлагалось ввести свой параметр всего 20 раз [2]. В связи с этим появилась сильная необходимость в исследованиях старых и разработке новых путей проверки статистических гипотез для малого количества образов в выборке.

Хи-квадрат критерий, предложенный к использованию Карлом Пирсоном в начале 20 столетия, на сегодняшний день является одним из стандартных и известных. На протяжении долгого периода времени его применение было актуальным и удовлетворяло заданным требованиям эффективности. Если представить его в виде искусственного эквивалентного нейрона, то он будет иметь вид:

$$X^2 = \sum_{i=1}^{k} \frac{\left[\frac{n_i}{m} - \left(P(x_{i+1}) - P(x_i)\right)\right]^2}{P(x_{i+1}) - P(x_i)},$$

где m – количество образов в выборке; k – число равных интервалов гистограммы; n_i – число опытов, попавших в i интервал гистограммы.

Однако в современных реалиях данная конструкция имеет большое значение ошибки на малом количестве образов. Причиной этому является наращивание погрешностей ранее проведённых расчетов. Порядка 30 % вероятности возникновения ошибок первого и второго рода при условии, что мы берем выборку, состоящую из 21 опыта и квантуем выходные данных по порогу 0.375. Согласно ГОСТ 52633.5–2011 [3] данная ошибка не должна превышать 3 %, в связи с этим хи-квадрат не подходит, как самостоятельный критерий для проверки закона нормального распределения.

Еще одним популярным, но менее известным является статистический критерий Андерсона-Дарлинга, открытый двумя учеными в 1952 году. Он, как и хи-квадрат критерий, имеет достаточную для практики мощность только при наличии выборки в 200 и более опытов, что также является его существенным недостатком. Его представление в виде искусственного эквивалентного нейрона будет выглядеть следующим образом:

$$AD^2 = \sum_{i=1}^{m} \frac{[i - m \cdot P(x_i)]^2}{[1 - P(x_i)] \cdot P(x_i)},$$

где m – количество образов в выборке.

При квантовании данных с выхода сумматора по порогу 232 и 21 количестве образов в выборке возникновение вероятности ошибок первого и второго рода будет равна 35 %, что на 5 % выше, чем у хи-квадрат критерия. Предложенный Пирсоном критерий оказывается несколько мощнее своего аналога [4].

За прошедший XX век было изучено и открыто порядка 45 статистических критерия: 21 для проверки гипотезы нормального распределения данных и 24 для проверки гипотезы равномерного распределения данных. Все они имеют различные вероятности появления ошибок первого и второго рода. Каждый из них можно представить в виде эквивалентного им искусственного

нейрона [5]. Также некоторые из них могут быть записаны в дифференциальной или логарифмической формах, что снижает вероятность появления ошибок как первого, так и второго рода. Например, семейство критериев Крамера-фон Мизера при 16 опытах имеет $P_1 \approx P_2 \approx 0.4$, а его дифференциальный вариант $P_1 \approx P_2 \approx 0.04$, что показывает нам снижение P_1 и P_2 в десять раз. Мы можем наблюдать это исходя из графиков.

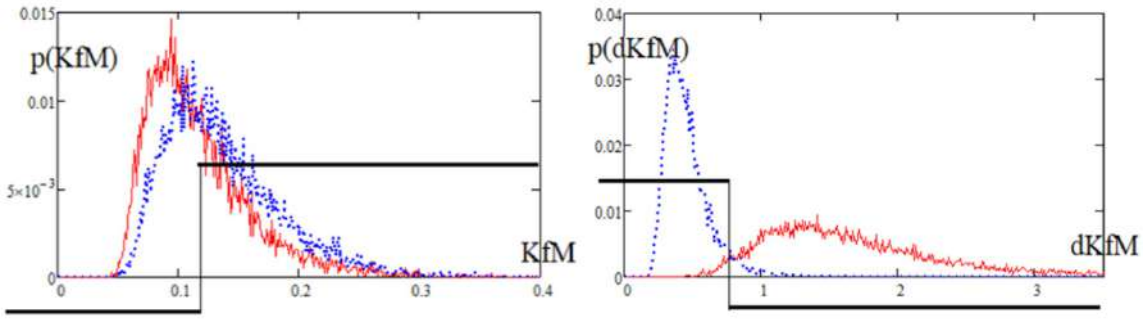


Рис. 1. Распределение данных классического (слева)
и дифференциального (справа) критерия Крамета-фон Мизера

Данный метод также действует и на различные модификации этого критерия, снижая вероятность появления ошибок первого и второго рода: Смирного ≈ в 8,4 раза, Андерсона-Дарлинга ≈ в 6,4 раза, Фроцини ≈ в 4,2 раза [6].

Для понижения P_1 и P_2 принято использовать сразу несколько статистических критериев, объединяя их в один слой нейронной сети. Данное решение позволяет сильно снизить вероятность появления ошибок первого и второго рода. Еще одним способом снижения данной вероятности может являться переход от нейросетевого анализа данных в статике к анализу в динамике. Одним из способов реализации данного перехода, может быть, применение модуляции входных данных, то есть изменение этих данные путем произвольного прореживания исходной выборки до более меньших подвыборок [7]. Рассмотрим это на примере 21 опыта в выборке и 15 опытов в подвыборке. Для данного случая число сочетаний будет ≈ 54264.

Таким образом, создать динамику изменения состояний выходного слоя нейронной сети можно воспользовавшись нейросетевым вычислителем модулятора данных на входе. Это позволит значительно увеличить правильность принимаемых решений.

Список литературы

1. Иванов А. И., Перфилов К. А., Малыгина Е. А. Многомерный статистический анализ качества биометрических данных на предельно малых выборках с использованием критериев среднего геометрического, вычисленного для анализируемых функций вероятности // Измерение. Мониторинг. Управление. Контроль. 2016. № 2 (16). С. 64–72.

2. Иванов А. И. Искусственные математические молекулы: повышение точности статистических оценок на малых выборках (программы на языке MathCAD). Пенза : Изд-во АО «Пензенский научно-исследовательский электротехнический институт» (АО «ПНИЭИ»), 2019. 32 с.

3. ГОСТ Р 52633.5–2011. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа. М. : Стандартинформ, 2011. 20 с.

4. 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.

5. Иванов А. П., Иванов А. И., Безяев А. В. [и др.]. Альбом статистических критериев, ориентированных на совместное использование при проверке гипотезы нормального или равномерного распределения данных малых выборок : препринт. Пенза, 2022. 22 с. doi: 10.13140/RG.2.2.15891.76324

6. Волчихин В. И., Иванов А. И., Перфилов К. А. [и др.]. Быстрый алгоритм обучения сетей искусственных нейронов квадрата среднего геометрического плотностей распределения значений многомерных биометрических данных // Известия высших учебных заведений. Поволжский регион. Технические науки. 2018. № 3. С. 23–35.

7. 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.

УДК 004.056

УСТРАНЕНИЕ ЭФФЕКТА ОШИБОЧНОГО НАБЛЮДЕНИЯ АНТИПЕРСИСТЕНТНОСТИ СВЯЗЕЙ ПРИ ОЦЕНКАХ ПОКАЗАТЕЛЯ ХЁРСТА НА МАЛЫХ ВЫБОРКАХ ЗА СЧЕТ ПОДБОРА ПОКАЗАТЕЛЯ ЛОГАРИФМИРОВАНИЯ

В. Е. Кузнецов[1], А. И. Иванов[2], В. Ю. Герасин[3]

[1] Научно-производственное предприятие «Рубин», г. Пенза
[2] Пензенский научно-исследовательский электротехнический институт, г. Пенза
[3] Финансовый университет при Правительстве Российской Федерации, г. Москва

Аннотация. Показано, что при ограниченных выборках процедуры вычисления показателя Хёрста могут приводить к ошибочному наблюдению антиперсистентности связей анализируемых данных. Предложено устранить причину ошибочности наблюдений за счет специальной балансировки объема выборки и показателя используемых при вычислениях логарифмов. Дана таблица значения показателей логарифмов для выборок в 16, 32, …, 2048 опытов, характерных для анализа биометрических данных и статистических данных колебаний рынка. Регуляризация вычислений выполнена для псевдослучайных генераторов независимых данных с нормальным законом распределения значений. Приводится кривая связи показателя логарифмов с размерами выборки.

Keywords: анализ данных биометрии, анализ данных рынка, малые выборки, регуляризация вычислений показателя Хёрста

ELIMINATING THE EFFECT OF ERROROUS OBSERVATION OF ANTI-PERSISTENCE OF CONNECTIONS WHEN ESTIMATING THE HURST INDICATOR ON SMALL SAMPLES DUE TO SELECTION OF THE LOGARITHMATION INDICATOR

V. E. Kuznetsov[1], A. I. Ivanov[2], V. Yu. Gerasin[3]

[1] Scientific-industrial Enterprise «Rubin», Penza
[2] Penza Scientific Research Electrotechnical Institute, Penza
[3] Financial University under the Government of the Russian Federation, Moscow

Abstract. It is shown that with limited samples, procedures for calculating the Hurst exponent can lead to erroneous observations of antipersistence of relationships

in the analyzed data. It is proposed to eliminate the cause of erroneous observations through special balancing of the sample size and the indicator of logarithms used in calculations. A table is given of the values of logarithm indicators for samples of 16, 32, ..., 2048 experiments, typical for the analysis of biometric data and statistical data of market fluctuations. Regularization of calculations was performed for pseudo-random generators of independent data with a normal distribution of values. A curve showing the relationship between logarithms and sample sizes is presented.

Keywords: biometric data analysis, market data analysis, small samples, regularization of Hurst exponent calculations

Вычисление показателей Хёрста в биометрии и экономике

Многие природные процессы описываются фрактальными связями, в том числе это относится к рынкам [1, 2] и к биометрическим данным [3]. В этом отношении усилия по созданию перспективных нейросетевых вычислителей [4] для приложений биометрии и приложений экономики могут быть объединены.

Одним из самых простых и в то же время самых очевидных фрактальных показателей является показатель Хёрста [1, 2], который оценивается как логарифм отношения размаха данных к их стандартному отклонению. К сожалению, подобные вычисления обладают плохой обусловленностью и как следствие выполняются на больших выборках в 1000 и более опытов. Для статистического анализа индексов рынка за несколько лет это вполне допустимо, однако для биометрии это недопустимо. Модели нейросетевой биометрии при обучении по ГОСТ Р 52633.5–2011 требуют примерно 16 примеров. То же самое должно возникать и при обучении нейросетевых моделей экономики [3].

Таким образом, возникает задача выполнять достаточно достоверные оценки показателей Хёрста на малых выборках в 16 опытов при создании нейросетевых приложений биометрии и экономики.

Пример программной реализации вычисления показателей Хёрста на малых выборках в 16 опытов для псевдослучайных (независимых) данных с нормальным распределением приведен в левой части рис. 1.

В правой части рисунка 1 дано распределение значений показателя Хёрста для выборки в 16 опытов. При росте объема выборки происходит нормализация распределения значений, что упрощает интерпретацию состояний исследуемой модели. Одновременно происходит рост математического ожидания значений показателя Хёрста.

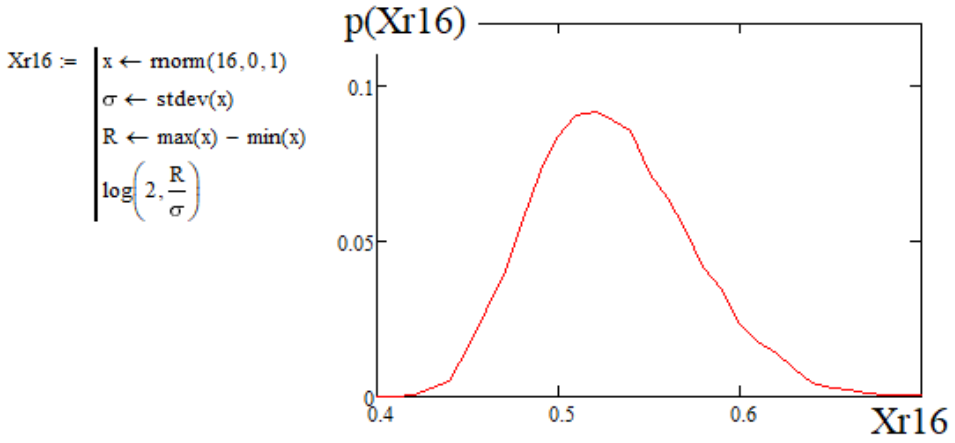


Рис. 1. Распределение значений показателя Хёрста для выборок в 16 опытов, полученных от программного генератора

Сам Хёрст пытался устранить эффект роста математического ожидания нормированием. При выборе логарифмов по основанию 2 мы получим:

$$Xr(x,N) = \frac{log\left\{2, \frac{max(x) - min(x)}{\sigma(x)}\right\}}{log\left\{2, \frac{N}{2)}\right\}} \qquad (1)$$

К сожалению, нормирование, использованное Хёрстом (1), приводит к усложнению физической интерпретации этого важного статистического показателя.

Оценка фрактальной размерности по шкале показателей Хёрста

Весьма интересным является подход к интерпретации, показателя Хёрста основанной на использовании генераторов псевдослучайных чисел [1, 2]. Исходя из результатов численных экспериментов, было доказано, что генератор независимых нормальных данных должен давать значение математического ожидания показателя Хёрста –0.5. При этом вопрос о выборе объема, анализируемой выборки остался открытым.

Для выборки в 16 опытов (рис. 1) математического ожидания показателей Хёрста, оказывается завышенным и составляет E(Xr16) = 0.535, что на 0.035 выше предположений теории фрактальной размерности.

Если увеличить объем выборки до 32 опытов, то математическое ожидание распределение Хёрста снижается до значения

E(Xr32) = 0.485, то так же не соответствует теории фрактальных размерностей.

Ситуация меняется, если для выборок в 16 опытов принять показатель логарифма – 1.91, а для выборок в 32 опыта принять показатель логарифма – 2.05. На рис. 2 приведен график, связывающий значения показателя логарифмирования в формуле Хёрста со значением, объема анализируемой выборки.

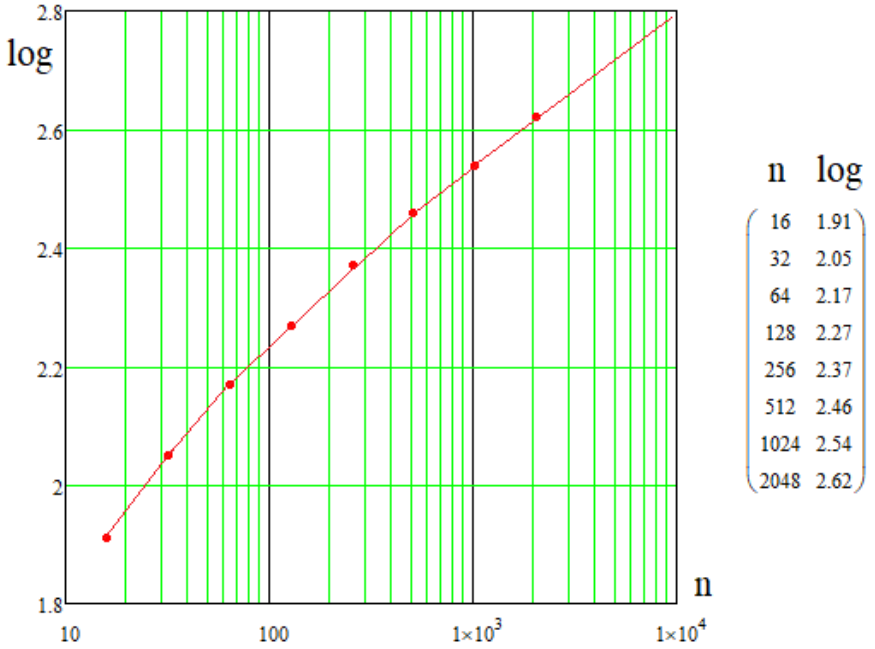


Рис. 2. Почти линейная связь оснований логарифмов с логарифмами размеров, анализируемой выборки

Проведенные исследования показали, что нормирование эмпирической формулы Хёрста (1) работает значительно хуже нормирования, выполненное выбором соответствующего показателя логарифмирования. Это важно при попытках физической интерпретации данных нейросетевой модели. Если показатель Хёрста меньше 0.5, то связи моделируемых данных антиперсистентны (в них утрачена низкочастотная составляющая шума, подавленная из-за завышенного интервала между отсчетами выборки).

Список литературы

1. Петерс Э. Хаос и порядок на рынках капитала. Новый аналитический взгляд на циклы, цены и изменчивость рынка / пер. с англ. В. И. Гусевой. М. : МИР, 2000. 333 с. ISBN 5-03-003356-4 (рус.), ISBN0-471-13938-6 (англ.)

2. Мандельброт Б., Хадсон З. Л. (НЕ)послушные рынки. Фрактальная революция в финансах. М. ; СПб. ; Киев : Вильямс, 2006. 408 с. ISBN 5-8459-0922-8, 0-1300-9717-9

3. Иванов А. И. Высокоразмерная коллективная биометрия подсознательного поведения людей на рынке и производстве : препринт // Пенза : Изд-во ПГУ, 2021. 60 с. ISBN 978-5-907456-44-0

4. Иванов А. И. Нейросетевой многокритериальный статистический анализ малых выборок : справочник. Пенза : Изд-во ПГУ, 2022. 160 с. ISBN 978-5-907600-83-6

УДК 004.056

БЫСТРЫЙ АЛГОРИТМ ВЫЧИСЛЕНИЯ ЭНТРОПИИ ДЛИННЫХ ЧИСЕЛ С ЗАВИСИМЫМИ РАЗРЯДАМИ

Н. И. Серикова[1], А. И. Иванов[2],
С. В. Куликов[3], А. Ю. Малыгин[4]

[1] Научно-производственное предприятие «Рубин», г. Пенза
[2,3] Пензенский научно-исследовательский электротехнический институт, г. Пенза
[4] Пензенский государственный университет, г. Пенза

Аннотация. Вычисление энтропии по Шеннону является задачей с экспоненциальной вычислительной сложностью. Предложено упростить задачу за счет симметризации матриц корреляционных связей с одинаковыми коэффициентами корреляции вне диагонали. Для симметричных матриц аналитическая связь энтропии с коэффициентом корреляционной сцепленности является квадратичной. В реальных условиях предложено оценивать коэффициент корреляционной сцепленности эквивалентной симметризованной матрицы усреднением модулей ее коэффициентов корреляции, находящихся вне диагонали.

Ключевые слова: энтропия по Шеннону, корреляционная сцепленность

FAST ALGORITHM FOR CALCULATING THE ENTROPY OF LONG NUMBERS WITH DEPENDENT BITS

N.I. Serikova[1], A. I. Ivanov[2],
S. V. Kulikov[3], A. Yu. Malygin[4]

[1] Scientific-industrial Enterprise «Rubin», Penza
[2,3] Penza Scientific Research Electrotechnical Institute, Penza
[4]Penza State University, Penza

Abstract. Calculating Shannon entropy is a problem with exponential computational complexity. It is proposed to simplify the problem by symmetrizing the correlation matrices with identical correlation coefficients outside the diagonal. For symmetric matrices, the analytical relationship between entropy and the coefficient of correlation entanglement is quadratic. In real conditions, it is proposed to estimate the coefficient of correlation entanglement of an equivalent symmetrized matrix by averaging the modules of its correlation coefficients located outside the diagonal.

Keywords: Shannon entropy, correlation entanglement

Введение

Задача вычисления энтропии длинных кодов по Шеннону в общем случае имеет экспоненциальную вычислительную сложность и требует для вычислений больших объемов исходных данных. То есть для вычисления энтропии одного разряда бинарного кода H(«x_1») достаточно знать порядка 100 состояний. Требуется вычислить вероятность появления состояний «0» и «1». Оценка двухмерной энтропии H(«x_1, x_2») является двухмерной задачей и для ее решения потребуется порядка 10 000 отсчетов четырех состояний «00», «01», «10», «11». Этих данных будет также достаточно для вычисления коэффициента парной корреляции r(«x_1, x_2»).

Очевидно, что рост размерности задачи быстро приводит к тому, что вычислительные ресурсы и ресурсы исходных данных быстро заканчиваются. Однако есть исключение, используемое обычно в криптографии. В случае, если все разряды длинных чисел независимы r(«x_1, x_2») = 0 = r(«x_1, x_3») = 0 = … = r(«x_{255}, x_{256}») = 0 и в каждом разряде равновероятны состояния «0» и «1», то энтропия легко оценивается. В этом случае, энтропия каждого разряда составляет 1 бит, а общая энтропия растет пропорционально длине разрядов кода. Для ключа длинной в 256 бит, энтропия составит 256 бит.

Двумя дополнительными точками вырождения сложных вычислений являются точки r(«x_i, x_j») = ±1. В этих двух случаях наблюдается полное вырождение случайной компоненты, мы наблюдаем только одномерную связь данных. Для нас важен факт того, что связь двухмерной энтропии с коэффициентом парной корреляции квадратична:

$$H("x_1, x_2") = 2 \cdot (1 - \{r("x_1, x_2")\}^2) \qquad (1)$$

При значениях коэффициента корреляции ±1 двухмерная энтропия оказывается нулевой. При нулевом значении коэффициента парной корреляции двухмерная энтропия имеет максимум H(«x_1, x_2») = 2.

Легко показать, что столь простая аналитическая связь (1) возникает из-за того, что двухмерная корреляционная матрица абсолютно симметрична. Вне ее диагонали лежат два одинаковых коэффициента корреляции. Если сохранять это свойство симметрии корреляционных матриц, то простота записи (1) будет сохраняться:

$$\begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & r & r \\ r & 1 & r \\ r & r & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & r & r & r \\ r & 1 & r & r \\ r & r & 1 & r \\ r & r & r & 1 \end{bmatrix}, \ldots \qquad (2)$$

Если добиться выполнения условия симметричности корреляционных матриц (2), то оценка энтропии случайных чисел любой длины будет иметь квадратичную вычислительную сложность:

$$\begin{cases} H("x_1, x_2, x_3") = 3 \cdot (1 - r^2), \\ H("x_1, x_2, x_3, x_4") = 4 \cdot (1 - r^2), \\ :: \\ H("x_1, x_2, \ldots.., \ x_{256}") = 256 \cdot (1 - r^2) \end{cases} \qquad (3)$$

Если корреляционная матрица полностью симметрична (совпадают все значения коэффициентов корреляции вне диагонали), то упрощается задача численного моделирования данных и оценки значения высокоразмерной энтропии.

Моделирование данных одинаково сцепленных друг с другом

Для того, чтобы программно воспроизвести одинаково сцепленные случайные данные необходимо воспользоваться симметричной связывающей матрицей. Достаточно получить от программного генератора псевдослучайных чисел вектор отсчетов и умножить его на симметричную связывающую матрицу [1, 2, 3]:

$$\begin{bmatrix} 1 & a & a & a & a \\ a & 1 & a & a & a \\ a & a & 1 & a & a \\ a & a & a & 1 & a \\ a & a & a & a & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{x}_3 \\ \tilde{x}_4 \\ \tilde{x}_5 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & \tilde{r} & \tilde{r} & \tilde{r} & \tilde{r} \\ \tilde{r} & 1 & \tilde{r} & \tilde{r} & \tilde{r} \\ \tilde{r} & \tilde{r} & 1 & \tilde{r} & \tilde{r} \\ \tilde{r} & \tilde{r} & \tilde{r} & 1 & \tilde{r} \\ \tilde{r} & \tilde{r} & \tilde{r} & \tilde{r} & 1 \end{bmatrix} \qquad (4)$$

В итоге мы получаем вектор одинаково связанных между собой случайных данных.

Пример программной реализации на языке MathCAD симметричного связывания приведен на рис. 1.

$$Mr(sh) := \left| \begin{array}{l} x \leftarrow \text{morm}(5, 0, 1 + sh) \\ a \leftarrow 0.42 \\ \text{for } i \in 0..4 \\ \quad y_i \leftarrow \left(\sum_{j=0}^{4} x_j \right) \cdot a + (1 - a) x_i \\ y \end{array} \right.$$

$i := 0..9999$

$rr_i := \text{rnd}(0.01)$ $\quad X^{\langle i \rangle} := Mr(rr_i)$

$b0_i := (X^{\langle i \rangle})_0$ $\quad b1_i := (X^{\langle i \rangle})_1$

$b2_i := (X^{\langle i \rangle})_2$ $\quad b3_i := (X^{\langle i \rangle})_3$

$b4_i := (X^{\langle i \rangle})_4$

corr(b0, b1) = 0.799

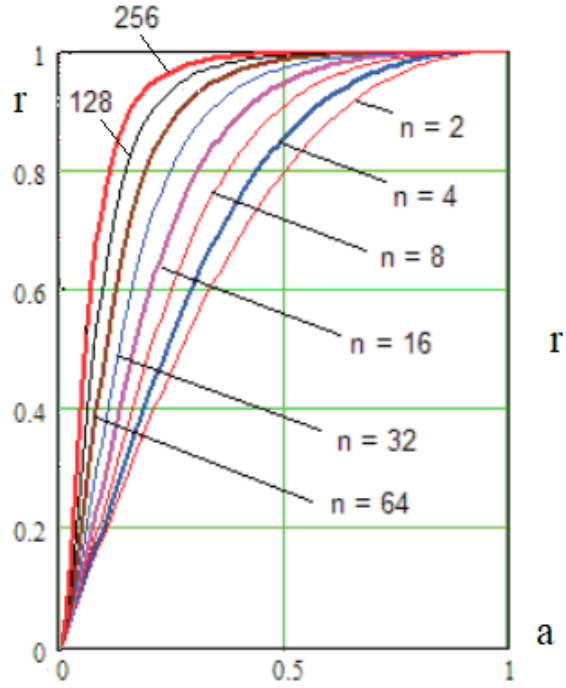
corr(b0, b2) = 0.801 corr(b1, b2) = 0.804

corr(b0, b3) = 0.796 corr(b1, b3) = 0.803 corr(b2, b3) = 0.799

corr(b0, b4) = 0.803 corr(b1, b4) = 0.805 corr(b2, b4) = 0.801

Рис. 1. Связывание данных без использования операций с матрицами

Для каждой размерности -n функции связывания коэффициентов равной корреляции и параметра связывания разные. Номограмма кривых связывания данных и аналитическая связь между настраиваемыми параметрами приведены на рис. 2.



$$r = \frac{2a + (n-2)a^2}{1 + (n-1)a^2} \qquad (5)$$

Рис. 2. Номограмма связи значений коэффициентов связывающей матрицы с итоговыми коэффициентами равной коррелированности

Быстрое вычисление многомерной энтропии через симметризацию корреляционных матриц

На практике мы всегда имеем асимметричную корреляционную матрицу реальных данных. Простейшим методом перехода

от реальной асимметричной корреляционной матрицы к ее симметричному аналогу является усреднение модулей коэффициентов корреляции, лежащих вне диагонали:

$$\begin{bmatrix} 1 & r_{1,2} & r_{1,3} & r_{1,4} & r_{1,5} \\ r_{1,2} & 1 & r_{2,3} & r_{2,4} & r_{2,5} \\ r_{1,3} & r_{2,3} & 1 & r_{3,4} & r_{3,5} \\ r_{1,4} & r_{2,4} & r_{3,4} & 1 & r_{4,5} \\ r_{1,5} & r_{2,5} & r_{3,5} & r_{4,5} & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & \tilde{r} & \tilde{r} & \tilde{r} & \tilde{r} \\ \tilde{r} & 1 & \tilde{r} & \tilde{r} & \tilde{r} \\ \tilde{r} & \tilde{r} & 1 & \tilde{r} & \tilde{r} \\ \tilde{r} & \tilde{r} & \tilde{r} & 1 & \tilde{r} \\ \tilde{r} & \tilde{r} & \tilde{r} & \tilde{r} & 1 \end{bmatrix}, \quad где \quad \tilde{r} \approx E\{|r_{i,j}|\} - \quad (6)$$

Очевидно, что преобразование (6) является приближенным, однако его методическая ошибка быстро падает с ростом размерности решаемой задачи. Предположительно, что в ближайшем будущем, может быть, разработан программный корректор методических ошибок.

Еще один путь снижения методических погрешностей перехода к эквивалентным симметричным корреляционным матрицам является использование хи-квадрат преобразований [4].

Список литературы

1. Малыгин А. Ю., Волчихин В. И., Иванов А. И., Фунтиков В. А. Быстрые алгоритмы тестирования нейросетевых механизмов биометрико-криптографической защиты информации. Пенза : Изд-во ПГУ, 2006. 161 с.

2. Ахметов Б. С., Волчихин В. И., Иванов А. И., Малыгин А. Ю. Алгоритмы тестирования биометрико-нейросетевых механизмов защиты информации. Казахстан, Алматы : Изд-во КазНТУ им. Сатпаева, 2013. 152 с. ISBN 978-101-228-586-4

3. Ахметов Б. С., Надеев Д. Н., Фунтиков В. А. [и др.]. Оценка рисков высоконадежной биометрии : монография. Алматы : Изд-во КазНТУ им. К. И. Сатпаева, 2014. 108 с.

4. Иванов А. И., Полковникова С. А. Хи-квадрат-симметризация корреляционных связей, ориентированная на одиночные нейроны и их обучение на малых выборках // Информационно-управляющие, телекоммуникационные системы, средства поражения и их техническое обеспечение : сб. науч. ст. IV Всерос. межведомственной науч.-техн. конф. Пенза, 2022. С. 131–141.

УДК 004.056

ОЦЕНКА СЛОЖНОСТИ ВОССТАНОВЛЕНИЯ БИОМЕТРИЧЕСКОГО ОБРАЗА ЧЕЛОВЕКА ИЗ НЕЙРОСЕТЕВОГО КОНТЕЙНЕРА ПРЕОБРАЗОВАТЕЛЯ БИОМЕТРИЯ-КОД

А. В. Майоров[1], М. В. Секретов[2]

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

Аннотация. Оценивается вычислительная сложность атаки восстановления биометрического образа человека из нейросетевого контейнера. Рассматривается применение преобразователя биометрия-код в системах парольной аутентификации, выполняющих вектор биометрических параметров лица человека в длинный трудно запоминаемый код доступа из случайных символов.

Ключевые слова: парольная аутентификация, нейросетевой контейнер, атака восстановления биометрического образа

ASSESSMENT OF THE COMPLEXITY OF RESTORING THE BIOMETRIC IMAGE OF A HUMAN FROM THE NEURAL NETWORK CONTAINER OF THE BIOMETRICS-CODE CONVERTER

A. V. Mayorov[1], M. V. Secretov[2]

[1,2] Penza Scientific Research Electrotechnical Institute, Penza

Abstract. The computational complexity of an attack to restore a biometric image of a person from a neural network container is estimated. We consider the use of a biometric-code converter in password authentication systems that convert a vector of biometric parameters of a person’s face into a long, difficult-to-remember access code of random characters.

Keywords: password authentication, neural network container, biometric image recovery attack

Введение

В системах парольной аутентификации наиболее уязвимым местом являются пароли. Простые легко запоминающиеся пароли также легко «взламываются», а сложные пароли, состоящие

из длинной последовательности случайных буквенно-цифровых символов, обычному человеку крайне сложно запомнить.

Для решения проблемы запоминания длинных паролей и облегчения их ввода может использоваться отечественная технология преобразования персональных биометрических данных в код. В основе указанной технологии лежат нейросетевые и криптографические преобразования персональных биометрических данных человека (бПДн) и случайного кода доступа (пароля). Важным свойством этой технологии является отсутствие прямой связи между бПДн человека и паролем. Связь между бПДн и паролем реализуется с помощью нейросетевого контейнера. В соответствии с определением ГОСТ Р 52633.4–2011, он представляет собой структурированный блок данных, содержащий параметры работы нейросетевого преобразователя.

Данные, содержащиеся в нейросетевом контейнере, потенциально, могут быть использованы злоумышленником для снижения сложности задачи восстановления пароля и даже получения биометрического образа человека. В настоящей статье описывается наиболее вероятный вариант атаки злоумышленника на биометрию пользователя с использованием нейросетевого контейнера и дается оценка стойкости нейросетевого контейнера к такой атаке.

1. Типовая схема применения преобразователя биометрия-код в системе аутентификации

Будем считать, что ПБК реализован с учетом требований ГОСТ 52633.0 [1], обучен по ГОСТ 52633.5 [2], протестирован по ГОСТ 52633.3 [3]. Параметры его работы для конкретного пользователя сохраняются в нейросетевом контейнере, защищенном от попыток исследований согласно ТС 26.2.002–2020 [4]. Преобразователь биометрия-код предназначен для получения, заданного во время «обучения» выходного кода (длинного пароля) при подаче на его входы биометрических данных и/или данных любой другой природы, отвечающих требованиям ГОСТ Р 52633.4.

Схема работы ПБК с включенным механизмом защиты от попыток исследований приведена на рис. 1. По ней бПДн пользователя обрабатываются с помощью некоторого первичного фильтра (сверточной искусственной нейронной сети), что позволяет получить вектор биометрических признаков человека из данных, считываемых с помощью биометрического сканера. Вектор биометрических признаков подается на вход «широкой» сети ПБК,

обученного формировать заданный выходной код для примеров биометрического образа «Свой». При подаче произвольных данных на выходе будет сформирован выходной код, отличающийся от правильного кода приблизительно в половине разрядов.

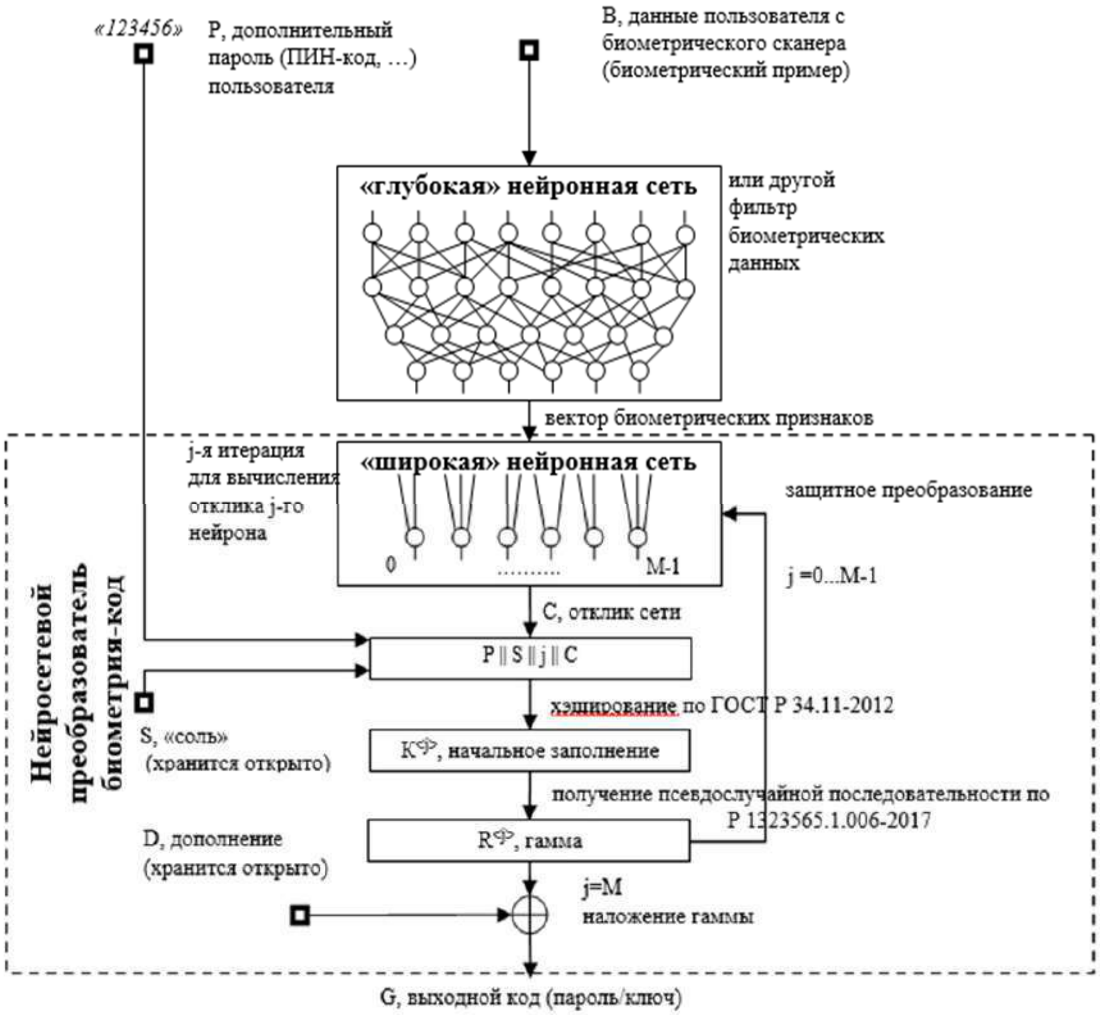


Рис. 1. Схема работы преобразователя биометрия-код

Для «обучения» ПБК во время регистрации пользователя выполняется считывание образцов его бПДн (например, лица), задается целевой выходной код (пароль) и проводится обучение «широкой» нейронной сети по правилам ГОСТ Р 52633.5 [2], а также защита параметров преобразования от попыток исследования согласно технической спецификации [4]. Пароль может быть выбран до начала ввода бПДн и, в общем случае, никак не связан с ними. Важно учесть, что для кодирования пароля необходимо выбирать такую кодировку, чтобы все варианты его битового представления были возможны и равновероятны.

После завершения обучения ПБК проводится тестирование, по результатам которого определяется достижение целевых показателей качества обучения. В случае если пользователь согласен с достигнутыми характеристиками, параметры преобразования сохраняются в виде нейросетевого контейнера, а исходные биометрические данные, производные от них биометрические признаки, пароль и другие компрометирующие пользователя данные удаляются.

ПБК имеет следующие особенности, отличающие его от других вариантов обработки биометрических данных:

1) позволяет формировать выходной код произвольной длины, который может интерпретироваться как пароль, личный ключ и т.д.;

2) не сохраняет биометрические данные пользователя, пароль и производные от них значения для работы;

3) хранимые параметры преобразования обладают стойкостью к попыткам исследования: 2^{L+M-1} попыток, где L – эффективная длина дополнительного пароля, подаваемого на вход нейросетевого преобразователя; M – число нейронов «широкой» ИНС, зависящее от качества предъявленных конкретным пользователем данных;

4) позволяет объединять несколько факторов аутентификации, связывая результат преобразования с некоторым дополнительным паролем, ПИН-кодом, идентификатором устройствам, несколькими видами биометрических данных;

5) не использует «решающее правило» или любые другие промежуточные значения, которые можно было бы использовать для оценки «правильности» вычисленного выходного кода (пароля);

6) настраивает структуру «широкой» ИНС уникальным образом во время каждой регистрации, даже для одних и тех же биометрических данных;

7) имеет возможность тестирования качества обучения «широкой» ИНС.

Рассмотрим типовую схему применения нейросетевого преобразователя биометрия-код (ПБК) для аутентификации пользователей с помощью паролей (рис. 2). Предъявленные пользователем данные и дополнительный пароль используются ПБК для воспроизведения выходного кода (пароля). Поскольку корректность полученного пароля ПБК проверить не может, он передает это значение дальше. На основе пароля вычисляется производное

значение, например, путем многократного хеширования с солью и некоторым случайным значением, созданным системой аутентификации. Результат сравнивается с ожидаемым контрольным значением, которое хранится в базе системы аутентификации, и принимается решение «Свой»/«Чужой». Число допустимых подряд идущих попыток аутентификации задается политикой безопасности системы аутентификации. Важно отметить, что контрольное значение не связано с биометрическими данными, а является производным от пароля.

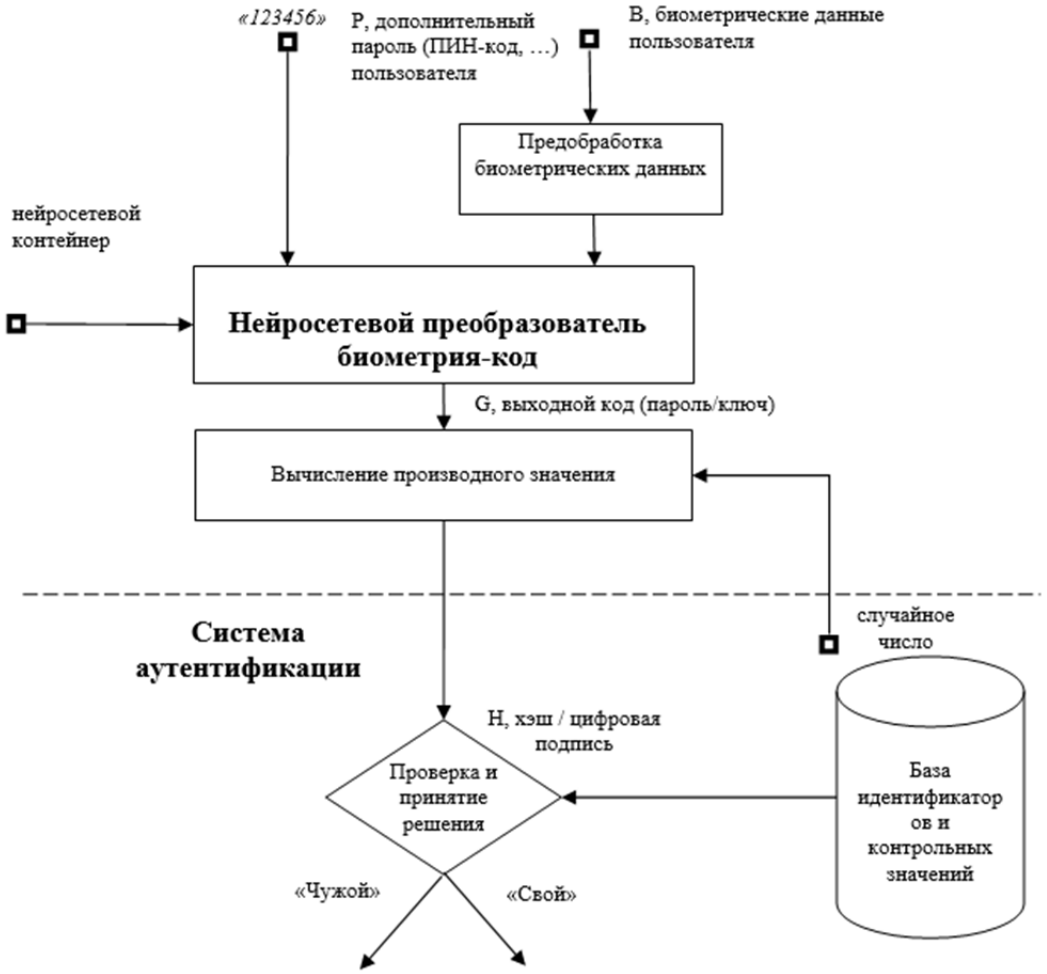


Рис. 2. Схема использования ПБК
в системе парольной аутентификации

Преобразование биометрия-код может выполняться как на сервере системы аутентификации, так и на клиентском устройстве. В последнем случае нейросетевой контейнер загружается на сторону клиента (на смартфон пользователя), используется для восстановления пароля, а затем производное от него значение возвращается на сервер для проверки. Нейросетевой контейнер

может формироваться и храниться на клиентском устройстве, что снимает проблему его хранения на сервере.

2. Хранимые данные в системе аутентификации, использующей ПБК

Для оценки безопасности предложенного решения следует перечислить хранимые в системе аутентификации данные и оценить возможность их использования злоумышленником для проведения атак.

К данным, общим для всех пользователей, можно отнести параметры алгоритмов (фильтров), используемых для обработки исходных биометрических данных пользователей, в том числе параметры искусственных нейронных сетей. Можно считать, что обучение ИНС выполнялось на базах биометрических данных лиц, которые никогда не были и не будут зарегистрированы в системе (биометрические образы/примеры «Чужие»). Поэтому, согласно ГОСТ ISO/IEC 2382–37 [5], эти параметры не компрометируют бПДн пользователя, хотя и могут потребоваться злоумышленнику для проведения атак.

К уникальным данных каждого пользователя можно отнести:

1) идентификатор пользователя (логин) или производное от него значение;

2) контрольное значение, полученное на основе пароля;

3) биометрический контейнер, если он хранится на сервере.

Идентификатор в простом случае представляет собой числовое значение, уникальное для каждого зарегистрированного пользователя в системе. Может быть порядковым номером пользователя в этой системе. Идентификаторы активно применяются в процессе обезличивания ПДн и сами по себе не позволяют установить личность субъекта ПДн.

Контрольное значение представляет собой значение, используемое для проверки знания/владения пользователем некоторого секрета (пароля, личного ключа). Злоумышленник может использовать это значение для подбора секрета пользователя, поэтому, в качестве контрольного значения обычно хранят результат многократного хэширования пароля/ключа с солью.

Нейросетевой контейнер содержит параметры нейросетевого преобразования в форме, гарантирующей его защиту от попыток исследования [4]. Контейнер включает в себя в защищенном виде таблицу связей, таблицу весов, соль и число-дополнение,

позволяющие правильно восстановить пароль из нейросетевого контейнера. Контрольные значения в контейнере не хранятся.

Если проводить аналогии, то нейросетевой контейнер представляет собой блок данных, который используется для восстановления паролей в типовых «хранителях паролей» (NordPass, Keeper, Enpass, 1Password, KeePassXC, …), только с возможностью восстановления защищенного пароля путем не ввода мастер-пароля, парольной фразы или графического пароля, а путем предъявления биометрических данных. Вместо нейросетевого контейнера пользователь может запоминать или хранить на физическом носителе длинный пароль и вводить его в систему аутентификации для подтверждения своего права доступа.

Таким образом, в системе парольной аутентификации, использующей нейросетевые преобразователи, ПДн и бПДн пользователя не хранятся, более того, в такой системе отсутствует «решающее правило» на основе биометрических данных. Поэтому, в отличие от типовых систем биометрической идентиификации/верификации, которые позволяют злоумышленнику после успешной атаки на сервер, получить прямой доступ к биометрическим данным пользователей или производным от них значениям [5], для реализации атаки на ПБК злоумышленнику придется дополнительно исследовать нейросетевой контейнер каждого пользователя.

3. Атаки на нейросетевой контейнер с целью восстановления биометрических данных

Оценим сложность атак восстановления изображения лица человека из нейросетевого контейнера: биометрические признаки лица человека – один из вариантов биометрической модальности, которые могут использоваться в ПБК. Будем считать, что при проведении атак злоумышленнику не известны бПДн пользователя, полученные из других источников, в противном случае, проводить атаку не требуется.

Атаку восстановления бПДн пользователя будем считать успешной, если злоумышленнику удалось восстановить исходное изображение лица пользователя и успешно использовать его в другой системе аутентификации. При этом будем рассматривать злоумышленников, в первую очередь внутренних, имеющих полный или частичный доступ к хранимым данным системы аутентификации. Злоумышленники, не имеющие доступ к хранимым

данным системы аутентификации, имеют гораздо меньше возможностей скомпрометировать даже существующие системы биометрической верификации/аутентификации.

3.1. Процедура восстановления изображения лица по вектору биометрических признаков

Заметим сразу, что если рассматривать типовую систему биометрической верификации/идентификации, то, в случае компрометации хранимых в ней записей (контрольных биометрических шаблонов, векторов биометрических признаков), эти данные могут использоваться злоумышленником сразу же для попыток синтеза бПДн пользователя.

Последние исследования искусственных нейронных сетей показывают, что, имея вектор биометрических признаков лица человека, полученный с помощью сверточной нейронной сети, можно не только частично восстановить биометрические характеристики (черты лица человека), но и синтезировать искусственное изображение реального лица с фотографическим качеством. Для решения этой задачи можно использовать генеративно-состязательные нейронные сети (GAN) [6–8]. Наиболее подходящим вариантом генеративно-состязательной сети, адаптированной для генерации высокодетализированных черт лица, является StyleGAN [9]. В проектах по генерации изображений лиц на основе 512 заданных биометрических признаков скрытого вектора синтезируют изображение лица человека, которое выглядит достаточно реалистично и близко к оригиналу.

Для восстановления изображения лица запускают цикл подбора, в котором, начиная с нулевых значений, случайно изменяются параметры скрытого вектора параметров StyleGAN для получения искусственного изображения лица. Для каждого сгенерированного лица вычисляют вектор биометрических признаков. От полученного вектора до известного вектора измеряют расстояние с помощью расстояния Евклида. Если расстояние между двумя векторами сократилось, то полученный вектор и расстояние запоминают как наиболее удачную попытку подбора. Действия повторяют множество раз, пока расстояние между векторами не станет меньше порогового значения, которое подбирают экспериментально. Для ускорения и уточнения работы алгоритма могут быть использованы заранее собранные базы лиц. Предварительное обучение StyleGAN, для которого необходимы мощные графические

ускорители, может и не потребоваться, так как в открытом доступе уже имеются предварительно обученные модели.

На рис. 3 показан результат восстановления изображения лица из его реального вектора биометрических признаков за 2000 итераций (около одного часа) на графическом ускорителе уровня Tesla P4 GPU. Видно, что достаточно хорошо удалось восстановить такие отличительные признаки человека как гендерную принадлежность, возраст, цвет и укладку волос. Для качественной работы StyleGAN требуется ее обучение на базе лиц реальных людей, в которой присутствуют люди разного возраста, пола, цвета кожи, этнической принадлежности.

Рис. 3. Пример восстановления изображения лица
из его вектора биометрических параметров (слева – реальное
изображение, справа – восстановленное из вектора
биометрических параметров)

В то же время собственные эксперименты показали, что реалистичное изображение лица восстановить не удается, если в обучающей базе StyleGAN не было близких типажей лица человека (присутствуют ошибки типа лица, цвета глаз). Дальнейшие исследования в этом направлении ведутся, в том числе, за счет эксплуатации информации внутренних связей слоев ИНС [13, 14].

Заметим также, что для осуществления подобной атаки злоумышленнику потребуется многократный доступ к сверточной ИНС, которая использовалась в системе аутентификации для получения вектора биометрических признаков. Если параметры работы ИНС (модель) недоступны злоумышленнику, то восстановить реальное изображение он не сможет.

3.2. Процедура восстановления пароля с помощью данных биометрического контейнера

Поскольку для систем аутентификации с ПБК биометрические вектора и шаблоны не хранятся, злоумышленнику выгодно пробовать атаку «с другой стороны», пытаясь сначала восстановить короткий пароль (например, из 8 символов), а затем, уже с его помощью подобрать длинный вектор биометрических признаков (например, 128 вещественных чисел) легального пользователя системы.

В системе аутентификации для проверки пользователя хранится некоторое контрольное значение. Для простоты рассуждений будем считать, что проверка производится по паролю (а не по личному ключу). Тогда контрольное значение будет представлять собой значение хэш-функции от этого пароля с солью. Для повышения вычислительной сложности и усложнения задачи восстановления пароля по хэш-значению, он может браться многократно. Будем считать, что для хэширования используется качественная криптографическая хэш-функция [12].

Злоумышленник, не имея доступ к хэш-значению, будет перебирать гипотезы значений пароля пользователя гарантированно долго, особенно, если система аутентификации имеет ограничение на число попыток в течение определенного промежутка времени. Поэтому будем считать, что, в худшем случае, злоумышленник имеет полный доступ к базе хэш-значений и может выполнять перебор на своем вычислительном устройстве. Тогда успешность подбора пароля будет зависеть от сложности пароля. В настоящее время рекомендуется использовать пароли не меньше 12 символов, содержащих символы алфавита в разных регистрах, цифры, тире, подчеркивание и другие знаки в случайном порядке.

С помощью ПБК можно перейти к паролям произвольной длины, например, 30-символьным. Этого достаточно для того, чтобы атака подбора пароля злоумышленником стала неосуществимой на практике, а все собранные им словари бесполезными. Понимая это, злоумышленник будет стремиться реализовать атаку с минимальным ожидаемым числом попыток. Будем считать, что злоумышленник получил доступ к нейросетевому контейнеру и использует хранимые в нем данные для снижения числа попыток подбора пароля.

Согласно результатам криптографических исследований, проведенных при разработке технической спецификации [4],

стойкость нейросетевого контейнера к попыткам исследования составляет 2^{M+L-1} попыток, где M – полученное в ходе обучения число нейронов выходного слоя «широкой» ИНС нейросетевого преобразователя; L – эффективная длина дополнительного пароля (ПИН-кода, пароля и другого секрета). Значение M зависит от используемого типа биометрических данных, их качества и способа предобработки.

Если во время формирования нейросетевого контейнера не использовался дополнительный фактор защиты, то злоумышленник должен будет перебрать 2^M вариантов откликов «широкой» нейронной сети нейросетевого преобразователя, среди которых гарантированно окажется один верный. Получив гипотезу откликов «широкой» нейронной сети, злоумышленник сможет вычислить предполагаемое значение пароля и проверить его значение по контрольному значению.

Важно заметить, что согласно схеме защиты [4] во время перебора злоумышленник обязан не просто перебирать значения откликов нейронов «широкой» ИНС («1» или «0»), но и вычислять хэш-значения для каждого нейрона, чтобы раскрыть настоящие номера связей, весовые коэффициенты и, как производное от них значение – выходной код преобразователя. В противном случае вероятность совпадения сформированного пароля с правильным паролем будет стремиться к 0 (близка к вероятности коллизии используемой хэш-функции). Поэтому для проверки одной гипотезы пароля, потребуется вычислить не меньше $2 \cdot M$ хэш-значений (одно хэш-значение для вычисления соли и одно для нахождения псевдослучайной последовательности). Поэтому для перебора всех вариантов потребуется вычислить $2 \cdot M \cdot 2^M$ хэш-значений. Злоумышленник может оптимизировать перебор, сохраняя ранее вычисленные варианты в памяти, во время обхода дерева гипотез. Тогда ему потребуется выполнить не меньше $2 \cdot (2^M + 2^{M-1})$ вычислений хэш-значений.

При использовании дополнительного секрета (дополнительного пароля, ПИН-кода, …) злоумышленник должен будет строить дерево гипотез для каждого варианта дополнительного секрета. То есть выполнить $2^L \cdot 2^M$ попыток для полного перебора всех вариантов. Вычислительная сложность каждой попытки будет зависеть от алгоритма формирования контрольного значения (хэша).

Чтобы сопоставить сложность атаки, оценим число вариантов качественного 12-символьного пароля, в котором отдельные символы выбираются случайно и независимо друг от друга. Будем

считать, что алфавит символа такого пароля состоит из 64 различных символов. Тогда энтропия пароля составит 6 бит на символ или 72 бита на весь пароль целиком. Тогда общее число вариантов составит 2^{72}, а среднее число попыток для угадывания 2^{71}.

Современные искусственные нейронные сети глубокого обучения могут формировать вектор с длиной от 128 до 512 биометрических признаков лица человека [13, 14]. Для таких входных данных ПБК способен «обучать» до 60 нейронов «широкой» нейронной сети, $M = 60$. Использование в качестве второго фактора защиты 4-значного ПИН-кода, дает дополнительно $L = \log_2(10000) \approx 13{,}3$. Поэтому общее число попыток подбора пароля путем исследования контейнера составит $2^{73,3}$, а среднее число попыток для угадывания $2^{72,3}$. Как видно, стойкость контейнера к подбору пароля, даже без учета большей вычислительной сложности одного нейросетевого преобразования, сопоставима со стойкостью прямого перебора 12-символьного пароля.

Использование любого другого дополнительного фактора: дополнительной биометрии, например, голоса или отпечатка пальца, 8-значного пароля, 18-значного идентификатора устройств(а), с которых(ого) разрешен доступ пользователю, дополнительного значения, хранимого для каждого контейнера на сервере отдельно от нейросетевых контейнеров в доверенной вычислительной среде, – способно повысить стойкость контейнера к атакам извлечения пароля до требуемого нормативными документами уровня.

Кроме того, благодаря тому, что веса и связи при обучении нейросетевого преобразователя выбираются случайным образом, а дополнительный пароль является секретом, злоумышленник не может использовать свой опыт атаки на один нейросетевой контейнер для атаки на следующий, лежащий рядом, контейнер другого пользователя. Получается, что в отличие от систем биометрической верификации/идентификации, в которых биометрические признаки (шаблоны) необходимо хранить открыто для явного сравнения, атака на группу из T нейросетевых контейнеров, потребует ровно в T раз больше ресурсов для реализации атаки.

3.3. Процедура восстановления биометрических признаков по найденному паролю

Для оценки стойкости нейросетевого контейнера к восстановлению биометрических признаков пользователя будем рассматривать **гипотетическую** ситуацию, когда злоумышленник,

выполнив 2^{M+L-1} попыток в какой-то из них смог подобрать пароль и проверить его правильность с помощью контрольного значения, хранимого на сервере. Одновременно с этим он правильно восстановил номера используемых биометрических признаков для каждого нейрона преобразователя, а также соответствующие им значения весовых коэффициентов.

Поскольку каждый нейрон использует уникальные номера связей, злоумышленник может исследовать каждый нейрон по отдельности, пытаясь подобрать значения связанных с отдельным нейроном признаков.

Для наглядности и простоты рассмотрим ситуацию, когда один нейрон имеет только 2 входа. В случае большего или меньшего числа входов у нейрона, общая логика проведения атаки сохранится. Будем считать, что злоумышленник знает квантованное значение бита нейрона («1» или «0») и имеет некоторую базу векторов биометрических признаков, описывающих множество «Все Чужие» (показано серым на рис. 4), статистические характеристики которого ему известны. Разделяющая плоскость в двухмерном случае представляет собой линию (ее отрезок показан на рис. 4 черным).

Поскольку в весовые коэффициенты для каждого нейрона нейросетевого контейнера нормализуются относительно большего значения, то, проводя атаку, злоумышленник сможет получить лишь их отношение, т.е. углы наклона разделяющей прямой к осям значений признаков (x_1, x_2). Сделав гипотезу значений отклика нейрона, злоумышленник может определить, с какой стороны от разделяющей линии лежит множество «Свой» используемых нейроном биометрических признаков, но конкретное их положение он не знает. Он может предположить, что множество «Свой» лежит внутри множества «ВсеЧужие». Кроме того, он может предположить, что множество «Свой» пересекает перпендикуляр, восстановленный от разделяющей линии, поскольку алгоритм обучения ГОСТ Р 52633.5 стремится минимизировать ошибку 1 рода. Однако, на практике, в том числе, из-за недостаточного размера обучающей выборки это не всегда так.

Тем не менее, злоумышленник может в качестве начального приближения взять значения весовых коэффициентов в качестве значений биометрических признаков ($x_1 = w_1$, $x_2 = w_2$). Сделаем это для всех нейронов нейросетевого преобразователя и оценим, насколько близко злоумышленник восстановит исходный биометрический образ.

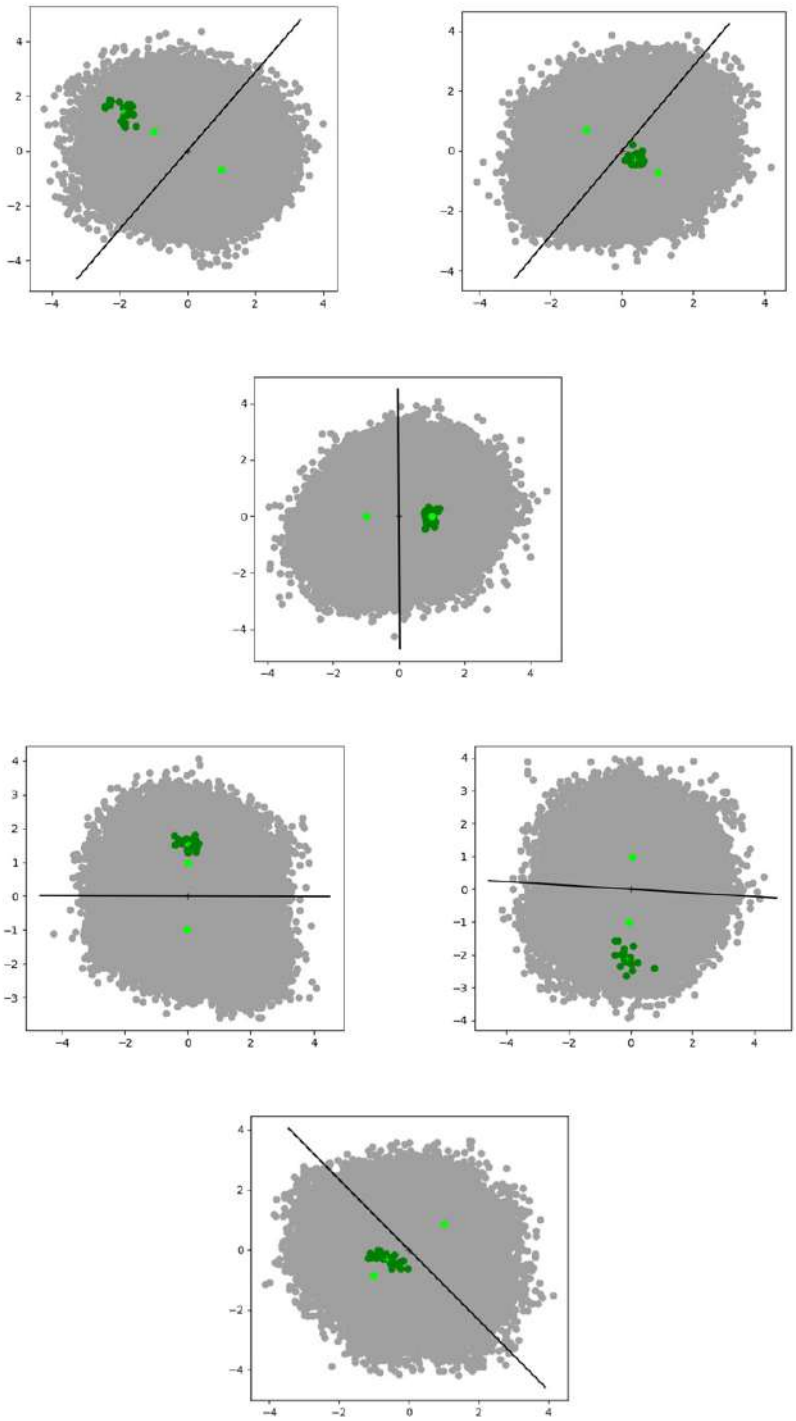


Рис. 4. Синтезированные значения биометрических признаков

На рис. 4 светло зеленым показана пара точек точки (w_1, w_2), (-w_1, -w_2), принятых злоумышленником в качестве гипотезы значений «Свой» и ее зеркального отражения. Как видно по рисунку реальные биометрические признаки могут лежать ближе к центру «Все Чужие», дальше от него, левее, правее, от восстановленных значений. Поэтому, злоумышленнику может потребоваться проверить 2, 3 или больше вариантов весовых коэффициентов, чтобы

подобрать правильное значение биометрических признаков. Для простоты проводимых оценок будем считать, что злоумышленник может найти значение неправильно восстановленного биометрического признака уже со второй попытки (хотя из рис. 4 следует, что это оптимистичный прогноз).

Будем считать, что для уверенности того, что восстановленные признаки являются принадлежащими пользователю, злоумышленнику нужно угадать значение биометрических признаков, лежащее в пределах ±3σ их реального распределения (считаем закон распределения биометрических признаков близким к нормальному). Тогда выбранные значения вектора биометрических признаков (x_1, x_2, ... x_N) будут входить в 99,7 % множества значений «Свой» и с большой вероятностью могут быть использованы в других биометрических системах. Значение N – число входных биометрических признаков.

Заметим, что если злоумышленник не угадает какой-то один биометрический признак, то ему придется делать одну или несколько дополнительных попыток угадывания этого признака в зависимости от его качества. Поскольку злоумышленник не знает, какой из биометрических признаков угадан неправильно, он должен перебрать все биометрические признаки, выполнив уже N повторных попыток. Если злоумышленник не угадает 2 биометрических признака, число попыток составит $N{\cdot}N\text{-}1$ и так далее. Таким образом, общее число попыток для числа биометрических признаков N и среднего числа не угаданных биометрических признаков K равно биномиальному коэффициенту C_N^K.

Оценка успешности восстановления значений биометрических признаков с помощью известных весовых коэффициентов проводилась экспериментально. Для этого оценивалась доля восстановленных значений биометрических признаков, попавших в ±3σ реальных биометрических образов «Свой». Эксперименты с нейросетевым преобразователем, обрабатывающим на входе вектор из 128 признаков лица человека, показали, что только около 79,8 % весовых коэффициентов были подобраны (восстановлены) правильно с первой попытки. Для оставшихся 20,2 % (25,6) должна быть выполнена как минимум вторая (третья) попытка. Поскольку злоумышленнику заранее не известно, в какой позиции они находятся, ему потребуется проверить не меньше C_{128}^{25} вариантов, что соответствует $2^{87,8}$ попыткам. На практике, злоумышленник может сократить число проверок, считая, что подбирать нужно только значащие биометрические признаки. Тогда ему удастся

сократить число вариантов до C_M^K, что соответствует C_{60}^{25} или $2^{55.5}$ попыток.

Заметим, что для проверки правильности каждого варианта злоумышленник должен восстановить лицо человека по способу, описанному в п. 3.1, либо передать его в некоторую классическую систему биометрической верификации/идентификации, которая позволит сравнить сформированный вектор с хранимым эталоном.

На рис. 5 показано распределение расстояний Евклида от синтезированного злоумышленником примера вектора биометрических признаков до реальных векторов «Свой» (зеленый), «Тест» (голубой), «Все Чужие» (серый). Из рисунка видно, что синтетический пример, хотя и находится ближе к примерам «Свой», чем к «Все Чужие», но не входит в их подмножество.

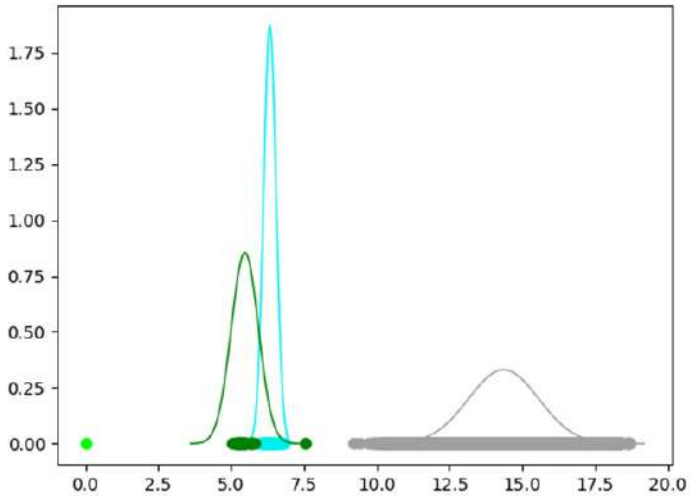


Рис. 5. Расстояние от синтезированного примера до обучающего и тестового, а также множества «Все Чужие»

Поскольку все системы биометрической верификации/идентификации используют разные критерии определения близости к образу «Свой» и правила настройки порога, по которому принимается решение «Свой»/«Чужой», ниже, на рис. 6 и 7 приведены результаты оценки близости с помощью расстояния Евклида и с помощью взвешенного расстояния Евклида соответственно. Во-втором случае результат оценки точнее, поскольку оценка близости вместо сферы проводится по гиперэллипсоиду, лучше учитывающего статистические особенности биометрического образа «Свой».

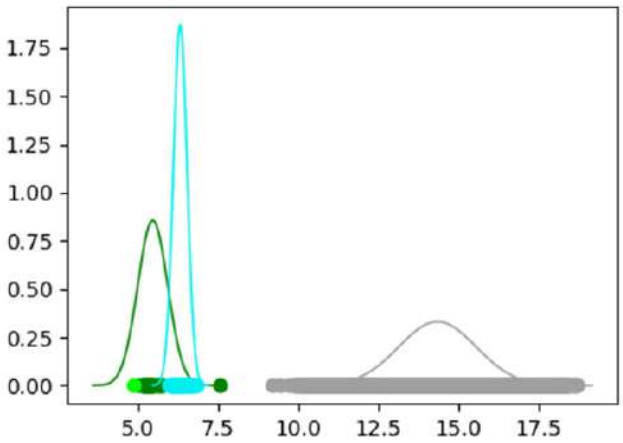


Рис. 6. Расстояние Евклида от центра «Свой»
до синтезированного примера «Синт»

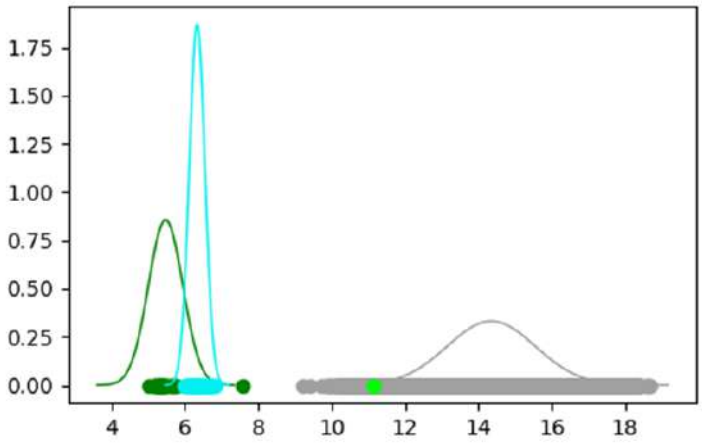


Рис. 7. Расстояние взвешенной меры Евклида от центра «Свой»
до синтезированного примера «Синт»

Как следует из рисунков, полученный синтетический пример может оказаться достаточным при использовании в простых системах идентификации, использующих евклидову меру для оценки образов друг к другу. Но, в случае, взвешенного евклидова расстояния, синтетический образ оказывается ничем не лучше других образов «Чужой».

Результаты восстановления лица человека по синтетическому вектору биометрических признаков, полученного путем анализа нейросетевого контейнера, показаны на рис. 8. Они подтверждают тот факт, что восстановленное изображение далеко от изображения реального человека, поэтому, на практике, не может использоваться повторно для «обмана» современных систем биометрической идентификации/верификации.

Рис. 8. Изображение реального лица человека (верхнее слева) и варианты его восстановления для одного и того же синтетического вектора биометрических признаков

Описанные эксперименты подтверждают, что восстановление биометрических данных человека для нейросетевого контейнера возможно только в случае полного и неограниченного доступа злоумышленника ко всем хранимым данным. При этом даже выполнение ресурсоемкого перебора всех вариантов не дает гарантий злоумышленнику успешного восстановления исходного лица пользователя. А в случае доведения стойкости нейросетевого контейнера к атакам до нормативного значения, обеспечивает надежную защиту от попыток его исследования.

Заключение

В работе показано, что ПБК, в отличие от систем биометрической идентификации/верификации не хранит личные биометрические данные человека или компрометирующие ее значения. Его применение создает для злоумышленника существенную (гарантированную) сложность восстановления биометрических данных на всех этапах проведения типовой атаки: подбора пароля, восстановления вектора биометрических признаков, синтеза биометрических данных (изображения лица человека). При этом злоумышленнику, для уверенности в успешной реализации атаки,

потребуется полный доступ к следующей информации: параметрам ИНС глубокого обучения, биометрическому контейнеру, контрольному хэш-значению. Отсутствие хотя бы одного из этих элементов делает атаку злоумышленника нереализуемой на практике.

По этим причинам преобразователи биометрия-код могут безопасно применяться в системах парольной аутентификации пользователей и не подпадают под действия п. 1 ст. 3 и ст. 11 Федерального закона от 27.07.2006 № 152-ФЗ «О персональных данных».

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 52633.3–2009. Защита информации. Техника защиты информации. Тестирование стойкости средств высоконадежной биометрической защиты к атакам подбора.

4. ТС 26.2.002–2020. Системы обработки информации. Криптографическая защита информации. Защита нейросетевых биометрических контейнеров с использованием криптографических алгоритмов.

5. ГОСТ ISO/IEC 2382-37–2016. Информационные технологии. СЛОВАРЬ. Ч а с т ь 37. Биометрия.

6. Chi Nhan Duong, Thanh-Dat Truong, Kha Gia Quach [et al.]. Vec2face: Unveil human faces from their blackbox features in face recognition, 2020.

7. Vishnu Naresh Boddeti. Secure face matching using fully homomorphic encryption. 2018.

8. Forrester Cole, David Belanger, Dilip Krishnan [et al.]. Synthesizing normalized faces from facial identity features. 2017.

9. Vendrow E., Vendrow J. Realistic face reconstruction from deep embeddings // in Proceedings of NeurIPS 2021 Workshop Privacy in Machine Learning. 2021.

10. Zhmoginov A., Sandler M. Inverting face embeddings with convolutional neural networks. arXiv preprint arXiv:1606.04189, 2016.

11. Zhigang Li and Yupin Luo. Generate identity-preserving faces by generative adversarial networks. 2017.

12. ГОСТ 34.11–2018. Информационная технология. Криптографическая защита информации. Функция хэширования.

13. Schroff F., Kalenichenko D., Philbin J. FaceNet: A Unified Embedding for Face Recognition and Clustering. Google Inc., 2015.

14. Akil Raihaniftee, Jakaria Rabbi, Mabrook S. Al-RakhamiRecent Advances in Deep LearningTechniques for Face Recognition // IEEE. 2021.

УДК 004.056

АППАРАТНО-ПРОГРАММНАЯ ЗАЩИТА НЕЙРОСЕТЕВОЙ БИОМЕТРИИ: СНИЖЕНИЕ ЭНЕРГОПОТРЕБЛЕНИЯ КОМПАКТНЫХ КОДОВ ПЕРЕБОРА С ОБНАРУЖЕНИЕМ И ИСПРАВЛЕНИЕМ ОШИБОК

А. В. Безяев[1], М. М. Бутаев[2], С. В. Качалин[3]

[1] Пензенский филиал Научно-технического центра «Атлас», г. Пенза
[2,3] Научно-производственное предприятие «Рубин», г. Пенза

Аннотация. Показано, что классические коды с обнаружением и исправлением ошибок большой избыточности плохо работают на «сырых» биометрических данных. Более эффективными являются самокорректирующиеся коды, построенные на вычислении хэш-функции и запоминании ее фрагментов. В работе предложено многократно ускорить перебор возможных хэш-состояний за счет замены вычислений криптографических функций на вычисление контрольных сумм CRC-32.

Ключевые слова: коды с обнаружением исправлением ошибок, биометрические данные, хэш-функция

HARDWARE AND SOFTWARE PROTECTION OF NEURAL NETWORK BIOMETRICS: REDUCING ENERGY CONSUMPTION OF COMPACT BRUX CODES WITH ERROR DETECTION AND CORRECTION

A. V. Bezyaev[1], M. M. Butaev[2], S. V. Kachalin[3]

[1] Penza branch of Scientific and Technical Center «Atlas», Penza
[2,3] Scientific-industrial Enterprise «Rubin», Penza

Abstract. It is shown that classical codes with detection and correction of errors of high redundancy do not work well on «raw» biometric data. More efficient are self-correcting codes based on calculating a hash function and storing its fragments. The work proposes to greatly speed up the search for possible hash states by replacing the calculation of cryptographic functions with the calculation of CRC-32 checksums.

Keywords: error detection codes, biometric data, hash function

Введение

В соответствии с Указом президента В. В. Путина № 490 от 10.10.19 «О развитии искусственного интеллекта в Российской Федерации» [1] к 2030 году Россия должна занять одно из ведущих

мест по практическому применению отечественных приложений ИИ. Во исполнение Указа [1] в 2019 году в России был создан технический комитет по стандартизации ТК 164 «Искусственный интеллект», который по его планам к 2024 году должен разработать и ввести в действие примерно 200 национальных стандартов.

К сожалению, далеко на любое приложение ИИ может быть пригодно к массовому использованию. В связи с этим в рамках национальной программы активно развиваются концептуальные требования к, так называемому, «Доверенному ИИ» [2]. На текущий момент под «Доверенным ИИ» понимаются приложения, выполняемые в доверенной вычислительной среде (SIM-карт, RFID-карт, микроSD-карт, USB-БиоТокенов, ПЛИС, DSP контроллеров). Как правило, контроллеры доверенной вычислительной среды имеет ограниченное число программируемых логических элементов, ограниченный объем ОЗУ и ПЗУ, что не позволяет применять алгоритмы высокой сложности для обработки данных.

В частности, в конце прошлого века зарубежные исследователи активно занимались использованием кодов с большой избыточностью при создании «нечетких экстракторов» [3, 4]. «Нечеткие экстракторы» выполняют квантование «сырых» биометрических данных, далее выполняют преобразование данных в избыточный код способный обнаруживать и исправлять ошибки. Это преобразование иллюстрирует рис. 1.

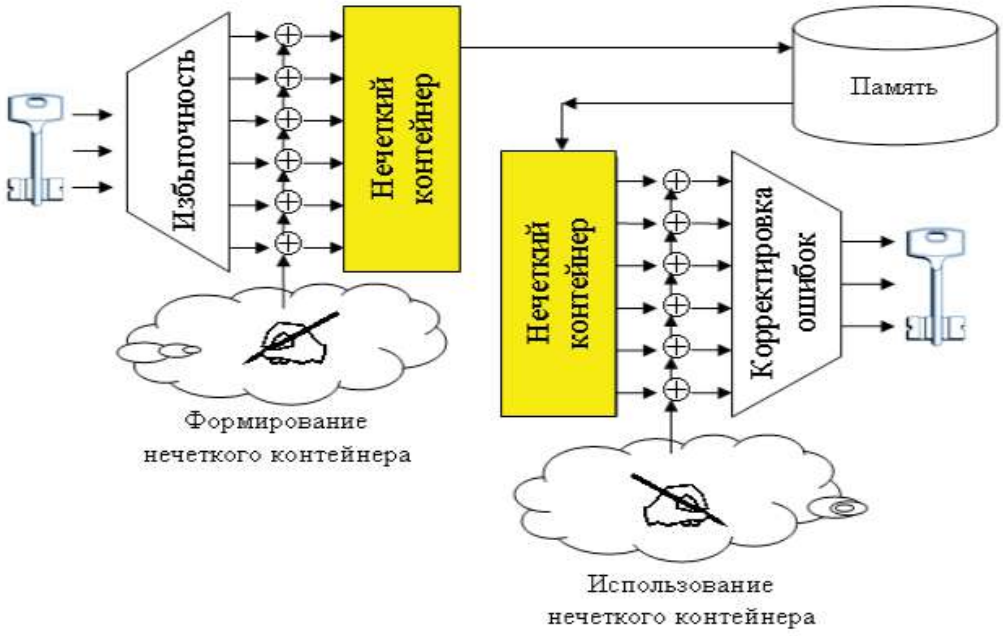


Рис. 1. Использование «нечеткими экстракторами» кодов для обнаружения и исправления ошибок в «сырых» биометрических данных

Из рис. 1 видно, что при формировании нечеткого контейнера создают короткий криптографический ключ. Далее вводят для него 20-ти кратную избыточность. Далее складывают по модулю два разряды код с высокой избыточностью с «сырыми» биометрическими данными. В итоге получают нечеткий контейнер «сырых» биометрических данных, который можно хранить открыто. При использовании нечеткого контейнера выполняют обратную операцию, получают длинный избыточный код. Затем код сворачивают, обнаруживая и корректируя в нем ошибки.

Основной проблемой «нечетких экстракторов» является малая длина ключа шифрования. Чем больше данных дает та или иная биометрическая технология, тем эффективнее оказываются «нечеткие экстракторы». В табл. 1 приведены данные по длине вектора «сырых» биометрических параметров, длине ключа нечеткого экстрактора, кодовой избыточности и вероятностям ошибок первого и второго рода.

Таблица 1

Данные по различным биометрическим технологиям

n	Биометрическая технология	Длина вектора БиоПарам	Длина ключа	Избыточность кода разы	P_1	P_2
1	Радужная оболочка глаза	2048	144	15	0.05	0.000001
2	Динамика рукописного почерка	416	22	19	0.05	0.0001
3	Отпечаток пальца	256	20	13	0.05	0.0001
4	Геометрия лица	128	11	12	0.05	0.001

Из табл. 1 следует, что первая биометрическая технологий (рисунок радужной оболочки глаза [3]) дает приемлемую длину криптографического ключа для протоколов биометрической аутентификации. Для других биометрических технологий длина криптографического ключа оказывается неприемлемо низкой [5].

Причиной нежелательного снижения длины криптографического ключа является то, что по классической схеме «нечетких экстракторов» (см. рис. 1) кодовая избыточность всегда приводит к утрате длины восстанавливаемого ключа.

Следует отметить, что этот эффект возможен только, если биометрические коды «сырые» и содержать до 30 % ошибок. Если же выполнить обогащение «сырых» биометрических данных, например, весовым суммированием, то положение кардинально

меняется. Чем выше уровень предварительного континуального обогащения [6], тем меньше требует кодовая избыточность самокорректирующегося классического кода.

При некотором числе входов у искусственных нейронов (персептронов) практически все ошибки биометрических данных исправляются при обучении нейросети алгоритмом ГОСТ Р 52633.5–2011. Тем не менее, при применении биометрии динамики рукописного почерка в условиях стресса уровень ошибок может повыситься и исправляющей способности сетей искусственных нейронов становится недостаточно. В этом случае могут быть использованы «не избыточные» самокорректирующиеся коды, построенные на хешировании фрагментов обогащенного выходного кода нейросети [7, 8, 9]. На рис. 2 приведена структура одной из реализаций такого типа кодов.

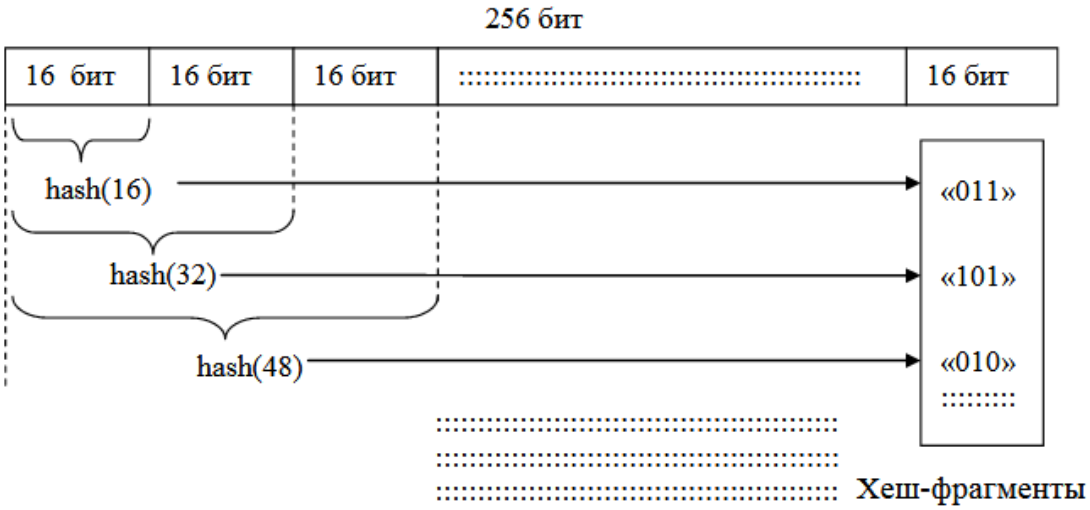


Рис. 2. Структура организации самокорректирующегося кода, построенного на хранении фрагментов хеш-функций выходного кода нейросети

В случае если мы будем строить самокорректирующийся код в рамках гипотезы появления по одной ошибке в одном из 16-ти битных фрагментов криптографического ключа, обнаружение мест ошибок может потребовать 16 х 16 = 256 кратное вычисление хеш-функций. При этом первые 16 хеш-функций следует вычислять над бинарными кодами в 16 бит, инвертируя в них по одному биту. Следующие 16 хеш-функций приходится вычислять над кодами длиной 32 бита, у которых должен инвертироваться один из последних 16 бит. При этом перебор возникает только в случае, если ошибка в 16-ти битном фрагменте кода действительно присутствует.

Как следствие, появление одной ошибки в последнем 16-ти битном фрагменте кода приведет к необходимости вычислять 15-ти хеш-функций над кодами растущей длины 16, 32, 48…, 240 бит. В случае, если мы используем в качестве доверенной вычислительной среды микросхему RFID-карты и пользуемся криптографическим хешированием, построенном на шифровании (блок в нижнем правом углу блок-схемы рис. 3) на каждое вычисление хеш-функции мы будем тратить одинаковое время (одинаковые энергоресурсы).

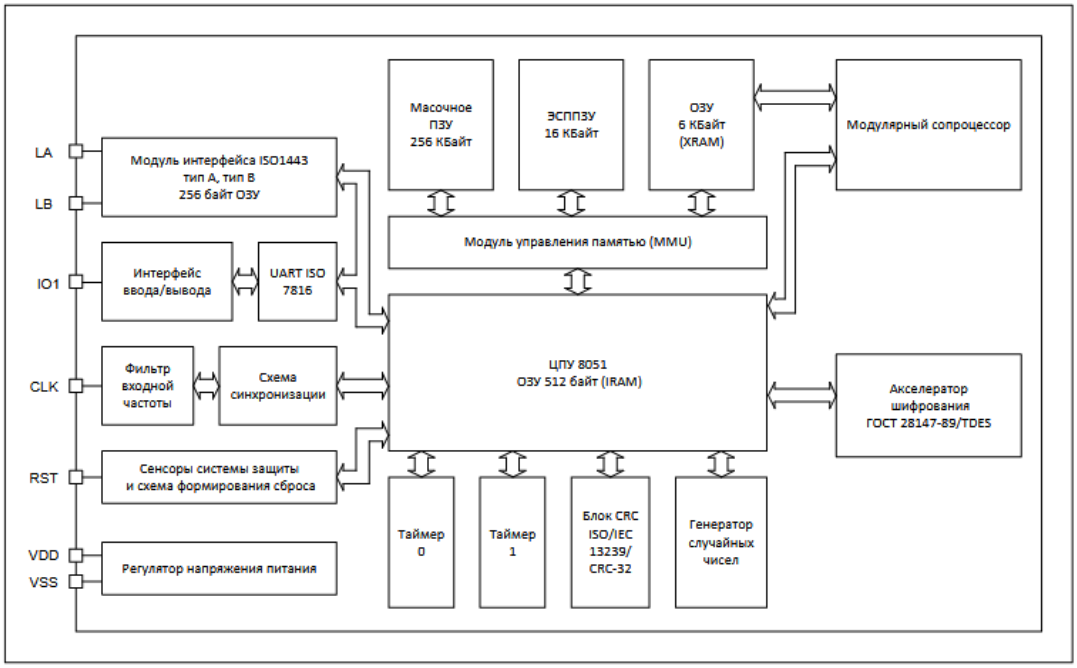


Рис. 3. Блок-схема контроллера микросхемы MIK51DC16D (R5016BKK01), способной работать как через контактную площадку, так и через RFID-карты

Сократить время вычисления хеш-функций примерно в 16 раза удается, если отказаться от использования криптографического хеширования, заменив его подсчетом контрольной суммы CRC-32 по международному стандарту ISO/IEC 13239. Подсчет контрольных сумм аппаратно реализован блоком, размещенным в центе нижней части рис. 3. При этом контрольную сумму целесообразно вычислять по двум примыкающим фрагментам кода по 16 бит. Первые 16 бит следует считать известной солью, а внутри следующих 16 бит следует выполнять перебор возможных значений кода через инвертирование одного бита.

Таким образом, более экономичной является структура самокорректирующегося кода, построенная на подсчете контрольных сумм, представленная на рис. 4.

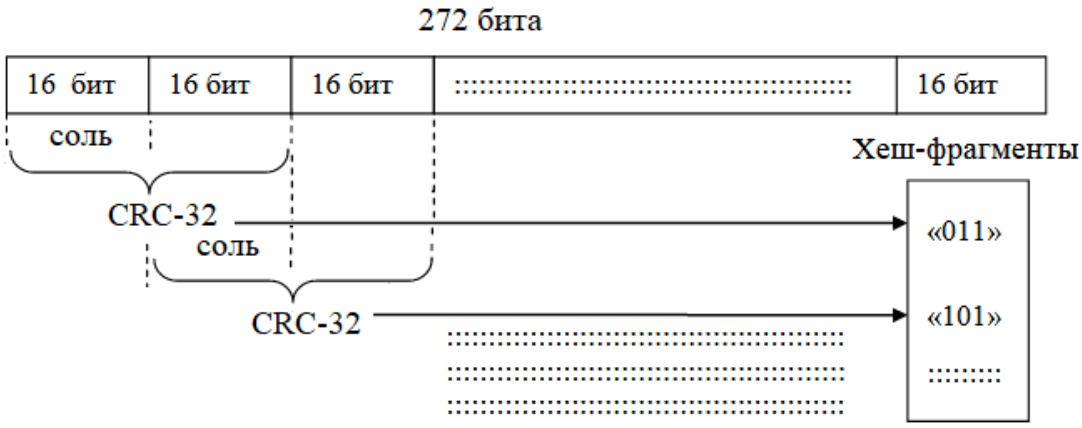


Рис. 4. Структура организации самокорректирующегося кода, построенного на хранении фрагментов не криптографической хеш-функций CRC-32 выходного кода нейросети

В итоге мы получаем самокорректирующийся код, имеющий компактную программную реализацию, ориентированную на использование ограниченных вычислительных ресурсов мало потребляющих доверенных контроллеров SIM-карт, RFID-карт, микроSD-карт, USB-БиоТокенов, ПЛИС микросхем, DSP-микросхем.

Список литературы

1. О развитии искусственного интеллекта в Российской Федерации : указ Президента РФ В. В. Путина № 490 от 10.10.19.

2. Цифровая экономика. Технологии доверенного искусственного интеллекта. 25 мая 2023. URL: https://trust-ai.ru

3. Dodis Y., Reyzin L., Smith A. Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy // EUROCRYPT. 2004. April 13. P. 523–540.

4. Чморра А. Л. Маскировка ключа с помощью биометрии // Проблемы передачи информации. 2011. № 2 (47). С. 128–143.

5. Безяев А. В., Иванов А. И., Ефимов О. В., Капитуров Н. В. Сравнение потенциальных возможностей классических и нейросетевых механизмов обнаружения и исправления ошибок, возникающих в биометрических кодах при аутентификации // Нейрокомпьютеры: разработка, применение. 2009. № 6. С. 41–44.

6. Иванов А. П., Кольчугина Е. А., Безяев А. В., Ерёменко Р. В. Снижение требований к корректирующей способности классических кодов с обнаружением и исправлением ошибок при использовании предварительного нейросетевого обогащения биометрических данных // Известия высших учебных заведений. Поволжский регион. Технические науки. 2022. № 3. С. 37–45.

7. Безяев, А. В. Нейросетевой преобразователь в самокорректирующийся код, совершенно не обладающий избыточностью // Нейрокомпьютеры: разработка, применение. 2012. № 3. С. 52–55.

8. Безяев А. В., Иванов А. И., Фунтикова Ю. В. Оптимизация структуры самокорректирующегося биокода, хранящего синдромы ошибок в виде фрагментов хеш-функций // Вестник Уральского федерального округа. Безопасность в информационной сфере. 2014. № 3 (13). С. 4–14.

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. ISBN 978-5-907262-59-1

УДК 004.056

МАЛЫЕ ВЫБОРКИ С НОРМАЛЬНЫМ И РАВНОМЕРНЫМ РАСПРЕДЕЛЕНИЕМ ДАННЫХ: НАСТРОЙКА ПАРАМЕТРОВ ДВУХ НЕПРЕРЫВНЫХ ВЫХОДНЫХ ФУНКЦИЙ ПРОГНОЗА ДОВЕРИЯ К РЕШЕНИЯМ ХИ-КВАДРАТ НЕЙРОНА

А. И. Иванов[1], Ю. И. Серикова[2], И. А. Филипов[3]

[1] Пензенский научно-исследовательский электротехнический институт, г. Пенза
[2] Научно-производственное предприятие «Рубин», г. Пенза
[3]Пензенский государственный университет, г. Пенза

Аннотация. Показано, что нейростетевой эквивалент хи-квадрат критерия при проверке гипотезы нормальности с выходным бинарным квантователем дает недостаточно информации. Предложено выполнять донастройку его выходных нелинейных функций с тем, что бы отклики нейрона дополнительно оценивали доверительную вероятность. Это позволяет строить более эффективные коды с обнаружением и исправлением ошибок.

Ключевые слова: малые выборки, хи-квадрат, коды с обнаружением и исправлением ошибок

SMALL SAMPLES WITH NORMAL AND UNIFORM DATA DISTRIBUTION: SETTING THE PARAMETERS OF TWO CONTINUOUS OUTPUT FUNCTIONS FOR PREDICTING CONFIDENCE IN CHI-SQUARE NEURON DECISIONS

A. I. Ivanov[1], Yu. I. Serikova[2], I. A. Filipov[3]

[1] Penza Scientific Research Electrotechnical Institute, Penza
[2] Scientific-industrial Enterprise «Rubin», Penza
[3]Penza State University, Penza

Abstract. It is shown that the neural network equivalent of the chi-square test when testing the normality hypothesis with an output binary quantizer does not provide enough information. It is proposed to perform additional adjustment of its output nonlinear functions so that the neuron responses would additionally evaluate the confidence probability. This allows you to build more efficient codes with error detection and correction.

Keywords: small samples, chi-square, error detection and correction codes

Введение

Известные итерационные алгоритмы обучения сетей искусственных нейронов [1, 2] неустойчивы из-за неустойчивости вычисления производных приращения либо снижения показателя качества обучения. Как следствие итерационные алгоритмы обучения трудно полностью автоматизировать. Проблему полной автоматизации обучения нейронов удается решить, если отказаться от итерационных процедур оптимизации и воспользоваться детерминированными процедурами приближенного вычисления весовых коэффициентов однослойной сети персептронов [3, 4].

При решении задач нейросетевой биометрико-криптографической аутентификации необходимо использовать доверенную вычислительную среду (SIM-карт, RFID-карт, микроSD-карт, USB-БиоТокенов, ПЛИС, DSP контроллеров). Как правило, доверенные контроллеры имеют ограниченные вычислительные ресурсы (ограниченный объем памяти, ограниченное число аппаратно реализованных криптографических функций, ограниченное энергопотребление). Предположительно, параллельно с аппаратной реализацией криптографических функций в доверенной вычислительной среде должны появиться аппаратно реализованные функции нейросетевых преобразований. В частности, перспективой является использование в контроллерах аппаратных реализаций искусственных нейронов являющихся аналогами статистических критериев [5, 6, 7], построенных на проверке гипотез о нормальном и равномерном распределении малых выборок.

Например, вектор из нескольких сотен биометрических параметров может быть использован для формирования малых выборок по 16 опытов с нормальным законом распределения значений и с равномерным распределением значений. Если воспользоваться классическим хи-квадрат критерием для проверки гипотезы нормального распределения данных, то мы получим распределения откликов, отображенное на рис. 1. В левой части рисунка дана программа, воспроизводящая критерий хи-квадрат при воздействии на него нормально распределенными данными.

В правой части рисунка отображены распределения откликов хи-квадрат критерия на нормальные и равномерные данные (для получения откликов на равномерные данные следует первую строку программы заменить на другую: x←sort(runif(16,-1,1)). Формально мы можем поставить в соответствие хи-квадрат критерию эквивалентный ему искусственный нейрон, поставив поле хи-квадрат обогатителя данных квантователь с порогом k = 5.9.

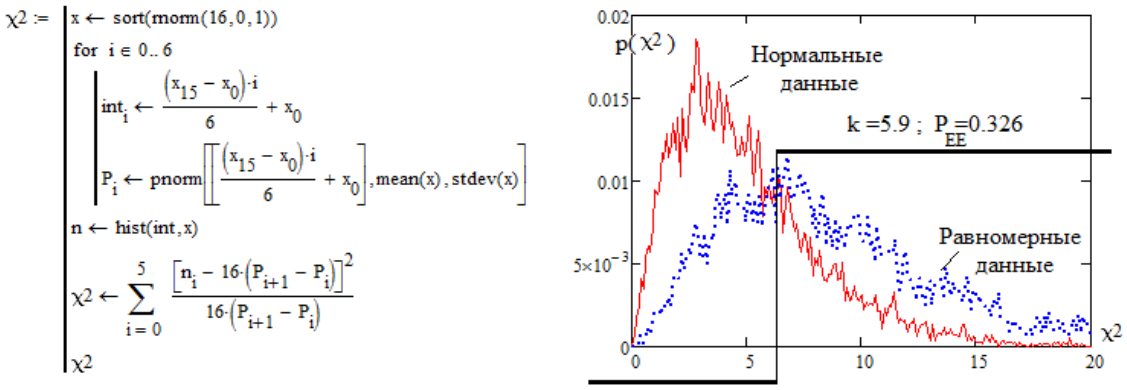


Рис. 1. Отклики хи-квадрат критерия настроенного проверять гипотезу нормального распределения малой выборки в 16 опытов

В этом случае искусственный нейрон будет откликаться состоянием «0» при обнаружении нормальных данных и состоянием «1» при обнаружении равномерно распределенных данных. Значение порога квантователя подобрано так, чтобы вероятности ошибок первого и второго рода искусственного нейрона совпадали P1 ≈ P2 ≈ PEE ≈ 0.326. То есть одиночный искусственный хи-квадрат нейрон, настроенный на проверку гипотезы нормальности позволяет принимать решения с низкой доверительной вероятностью 0.674.

Очевидно, что несколько статистических критериев, использованных параллельно, должны давать более достоверный результат в сравнении с одним критерием. Для их объединения могут быть использованы простейшие избыточные коды, способные обнаруживать и исправлять ошибки [7]. Очевидно, так же, что замена простейших кодов на более сложные конструкции должна так же приводить к росту достоверности принимаемых решений.

Дообучение искусственного хи-квадрат нейрона, введением двух дополнительных функций прогноза уровня доверия

В случае снижения обнаруженного значения хи-квадрат отклика растет доверие к выходному состоянию «0». При росте значения хи-квадрат увеличивается доверие к выходному состоянию «1». Рассуждая формально, мы можем параллельно с бинарным квантователем использовать две непрерывные монотонные выходные функции хи-квадрат нейрона, выходное состояние которых оценивает уровень доверия к дискретному решению искусственного нейрона. В итоге получается хи-квадрат нейрон с тремя выходами, отображенный на рис. 2.

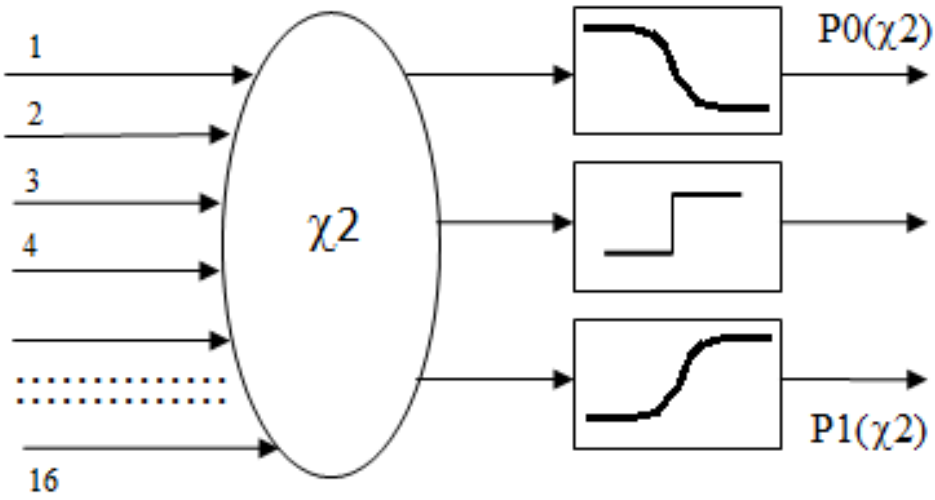


Рис. 2. Хи-квадрат нейрон с тремя выходами, два из которых непрерывные и дают оценку доверительной вероятности состояний третьего дискретного выхода

Очевидно, что дополнительные функции доверия P0(χ2) и P1(χ2) могут быть получены интегрированием плотностей распределения вероятностей, представленных на рис. 1. В частности, результаты интегрирования могут быть отображены таблицами доверительной вероятности, что является обычной формой для статистических справочников [8] и стандартов [9].

Ниже приведена таблица доверительных вероятностей, при необходимости на базе нее строится кусочно-линейная непрерывная функция оценок двух доверительных вероятностей.

Таблица доверительных вероятностей для хи-квадрат нейрона Пирсона -16

n	0	1	2	3	4	5	6	7	8	9	10
χ^2	0	1	3	5	7	8	9	10	12	15	30
P0	1	0.954	0.69	0.42	0.223	0.161	0.116	0.082	0.032	0.012	$1 \cdot 10^{-4}$
P1	0	0.01	0.094	0.253	0.437	0.53	0.603	0.674	0.789	0.894	0.997

Список литературы

1. Хайкин С. Нейронные сети: полный курс. М. : Вильямс, 2006. С. 1104.

2. Рассел С., Норвиг П. Искусственный интеллект. Современный подход. М. ; СПб. ; Киев, 2006. 1407 с.

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

4. Язов Ю. К., Волчихин В. И., Иванов А. И. [и др.]. Нейросетевая защита персональных биометрических данных / под ред. Ю. К. Язова. М. : Радиотехника, 2012. 157 с. ISBN 978-5-88070-044-8

5. 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. ISBN 978-5-907262-42-3

6. 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

7. Иванов А. И. Нейросетевой многокритериальный статистический анализ малых выборок. Справочник. Пенза : Изд-во ПГУ, 2022. 160 с. ISBN 978-5-907600-83-6

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

9. Р 50.1.037–2002. Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Часть I. Критерии типа $\chi 2$. Госстандарт России. М., 2001. 140 с.

УДК 004.67:519.7:57.087.1

ОЦЕНКА ЧИСЛА ПОТЕНЦИАЛЬНО РАЗДЕЛЯЕМЫХ КЛАССОВ КОРРЕЛЯЦИОННОЙ СЦЕПЛЕННОСТИ, НЕОБХОДИМЫХ ДЛЯ КОРРЕКТНОГО ОБУЧЕНИЯ КВАДРАТИЧНЫХ СЕТЕЙ ИСКУССТВЕННЫХ НЕЙРОНОВ

Т. А. Золотарева

Липецкий государственный педагогический университет имени П. П. Семенова-Тян-Шанского, г. Липецк

Аннотация. Показано, что существующие критерии проверки гипотезы независимости дают недостаточную достоверность принятия решений. Объединение нескольких критериев позволяет увеличить достоверность принимаемых решений. Предложен показатель, позволяющий оценить эффект от нейросетевого объединения критериев состоящий в увеличении объема обучающей выборки, обеспечивающей аналогичное обучение сетей квадратичных искусственных нейронов.

Ключевые слова: персептрон, коэффициент корреляции, корреляционная сцепленность

ESTIMATION OF THE NUMBER OF POTENTIALLY SEPARATED CLASSES OF CORRELATIONAL CONNECTION NECESSARY FOR CORRECT TRAINING OF QUADRATIC NETWORKS OF ARTIFICIAL NEURONS

T. A. Zolotareva

Lipetsk State Pedagogical University named after P. P. Semenov-Tyan-Shansky, Lipetsk

Abstract. It is shown that the existing criteria for testing the independence hypothesis provide insufficient reliability of decision making. Combining several criteria allows you to increase the reliability of decisions made. An indicator has been proposed that allows one to evaluate the effect of a neural network combination of criteria consisting in increasing the size of the training sample, which provides similar training for networks of quadratic artificial neurons.

Keywords: perceptron, correlation coefficient, correlation entanglement

Введение

Известно, что переход от использования сетей персептронов с накоплением данных в линейном пространстве к сетям квадратичных нейронов дает более высокое качество принимаемых

решений [1, 2]. Однако классический критерий Эджуорта-Эдлтона-Пирсона конца 19 века [3] на малых выборках в 16 опытов в место нулевой корреляции дает ошибки в интервале от –0.7 до +0.7. Эта ситуация отображена на рис. 1.

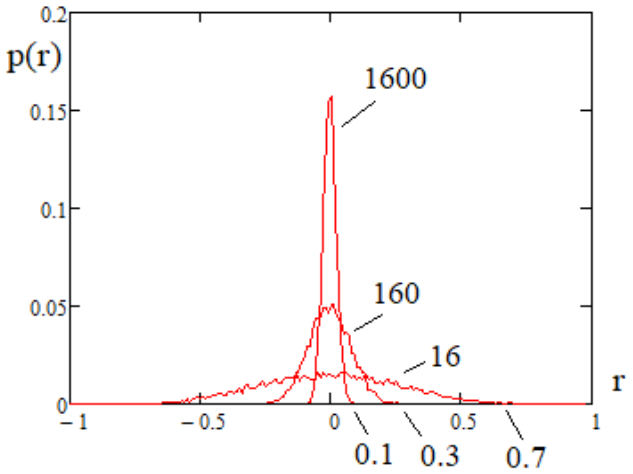


Рис. 1. Распределения значений ошибок оценки коэффициентов корреляции при вычислениях по классической формуле Эджуорта-Эдлтона-Пирсона конца XIX в. [3]

Из рис. 1 следует, что, опираясь только на классическую формулу вычисления коэффициентов корреляции на малых выборках в 16 опытов технически возможно оценивать коэффициенты корреляции 5 разных классов с шагом между центрами классов $\Delta r = 0.4$. На рисунке 2 представлены перекрывающиеся распределения пяти перекрывающихся классов коэффициентов корреляции.

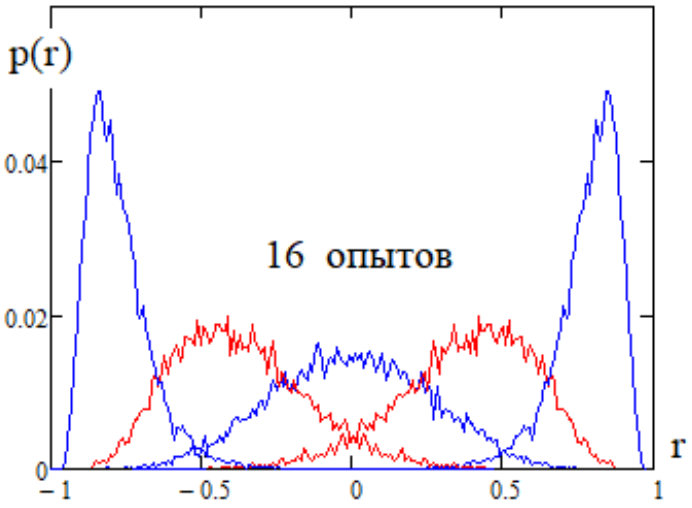


Рис. 2. Возможность создания дискретного (цифрового) вычислителя коэффициентов корреляции с шагом 0.4, ориентированного на малую выборку и классическую формулу Пирсона-Эджуорта-Эудлтона (16 входов у корреляционных нейронов)

Левый класс сильной отрицательной зависимости E(r) = –0.8; следующий класс меньшей корреляционной связанности E(r) = –0.4. Центральный класс слабо зависимых данных E(r) = 0.0 является самым широким и имеет пересечение со всеми другими классами корреляционной сцепленности. Следующий класс корреляционной сцепленности с центром E(r) = +0.4 пересекается с предыдущим классом на уровне равных вероятностей $P_{EE} \approx 0.3$. Последний класс сильно коррелированных данных с центром E(r) = +0.8 имеет существенно меньшее пересечение с предыдущим классом E(r) = +0.4.

Влияние числа входов у корреляционных нейронов на число разделяемых классов

Следует отметить, что отказ от возможного появления в выборке данных среднего уровня корреляционной сцепленности E(r) = ±0.4 многократно снижает уровень взаимного пересечения трех оставшихся классов корреляционной сцепленности до уровня $P_{EE} \approx 0.03$. Именно этот прием устранения взаимного пересечения разделяемых классов коэффициентов корреляции, использован в новом национальном стандарте России [4].

Очевидно, что 10-ти кратное увеличение числа входов у корреляционных нейронов стандарта [4] всегда будет приводить к росту числа разделяемых классов. Эта ситуация отображена на рис. 3.

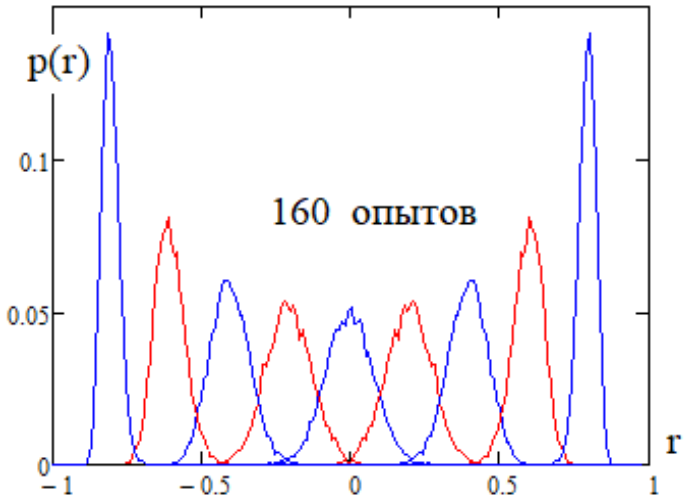


Рис. 3. Возможность создания достаточно точного дискретного (цифрового) вычислителя коэффициентов корреляции с шагом 0.2, ориентированного на большие выборки (160 входов у нейрона) и вычисления по классической формуле Пирсона-Эджуорта-Эудлтона

Из рис. 3 видно, что при использовании при реализации стандарта [4] корреляционных нейронов со 160 входами удается достаточно надежно разделять 9 классов данных с монотонно растущей корреляционной сцепленностью. При этом математические ожидания, разделяемых классов будут располагаться с шагом 0.2. E(r) = {–0.8, –0.6, –0.4, –0.2, –0.0, +0.2, +0.4, +0.6, +0.8}. Если мы при обучении сети корреляционных нейронов устраним 6 возможных классов, то получим три очень хорошо разделимых класса E(r) = {–0.8, –0.0, +0.8}.

Таким образом, мы всегда можем добиться разделения классов корреляционной сцепленности, увеличивая число входов у нейронов. Даже если вычисления выполняются даже по классической формуле Эджуорта-Эдлтона-Пирсона [3] всегда можно подобрать число входов у нейронов. Естественно, мы можем сократить число входов у нейронов за счет использования 2, 3, 4 и более параллельно используемых искусственных нейронов [5, 6]. Добавление числа искусственных нейронов должно снижать число входов у параллельно используемых нейронов. В этом плане появляется возможность использовать два пути регуляризации вычислений: регуляризацию можно выполнять либо за счет повышения числа входов у нейронов, либо за счет увеличения числа параллельно использовать несколько используемых корреляционных нейронов [5, 6].

Список литературы

1. 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.

2. Волчихин В. И., Иванов А. И., Малыгина Е. А., Серикова Ю. И. Сопоставление мощности 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. 47–57.

3. URL: https://ru.wikipedia.org/wiki/Корреляция

4. Проект стандарта «Искусственный интеллект. Нейросетевые алгоритмы в защищенном исполнении. Автоматическое обучение нейросетевых моделей на малых выборках в задачах классификации. URL: https://fgis.gost.ru

5. 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.

6. 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. ISBN 978-5-907364-24-0

УДК 004.056

АВТОКОРРЕЛЯЦИОННЫЙ КРИТЕРИЙ ОЦЕНКИ КОДОВЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ, ПОЛУЧЕННЫХ ИЗ НЕСТАБИЛЬНОЙ КОМПОНЕНТЫ БИОМЕТРИЧЕСКИХ ДАННЫХ, НА БЛИЗОСТЬ К «БЕЛОМУ» ШУМУ

А. В. Строков

ООО «Системы распределенного реестра», г. Москва

Аннотация. Предложено использовать автокорреляционную функцию для дискретных шумов по аналогии с классическим вычислением автокорреляционной функции шума с континуальными отсчетами. Почти «белый» шум предложено получать от программного генератора псевдослучайных чисел, образцы «окрашенного» шума предложено получать скользящей сверткой восьми рядом стоящих отсчетов «белого» шума без их взвешивания. Сумма модулей семи первых отсчетов автокорреляционной функции анализируемого шума является мощным критерием проверки гипотезы независимости дискретных данных выборкой в 256 бит. Этот критерий имеет низкую почти линейную вычислительную сложность и одновременно дает высокий уровень линейной разделимости зависимых и независимых данных. Мощность этого нового статистического критерия выше мощности аналогичных статистических критериев, построенных на вычислении расстояний Хэмминга.

Ключевые слова: «белый» шум, биометрические данные, статистический критерий оценки кодовых последовательностей, низкая вычислительная сложность

AUTO-CORRELATION CRITERION FOR ASSESSING CODE SEQUENCES OBTAINED FROM AN UNSTABLE COMPONENT OF BIOMETRIC DATA FOR PROXIMITY TO «WHITE» NOISE

A. V. Strokov

Distributed Registry Systems LLC, Moscow

Abstract. It is proposed to use the autocorrelation function for discrete noise by analogy with the classical calculation of the autocorrelation function of noise with continuous samples. It is proposed to obtain almost «white» noise from a software pseudo-random number generator; it is proposed to obtain samples of «colored» noise by sliding convolution of eight adjacent samples of «white» noise without weighing them. The sum of the moduli of the first seven samples of the autocorrelation function and the analyzed noise is a powerful criterion for testing

the hypothesis of independence of discrete data with a sample of 256 bits. This criterion has a low, almost linear computational complexity and at the same time gives a high level of linear separability of dependent and independent data. The power of this new statistical criterion is higher than the power of similar statistical tests based on the calculation of Hamming distances.

Keywords: «white» noise, biometric data, statistical criterion for evaluating code sequences, low computational complexity

Получение нестабильной компоненты биометрических данных

Средства биометрической защиты цифровых прав граждан Российской Федерации могут быть эффективны только в случае выполнения биометрических и криптографических преобразований в доверенной вычислительной среде, с обеспечением выполнения требований национальных стандартов, регламентирующих требования к нейросетевым преобразователям особенностей биометрии в кодовую последовательность (криптографический ключ) пользователя. Стоимость доверенной вычислительной среды должна быть низка, а ее применение должно быть массовым. Например, в качестве доверенной вычислительной среды могут быть использованы процессоры RFID идентификационных карт, SIM карт, micro-SD карт.

Одной из важнейших функций такой доверенной вычислительной среды является поддержка создания кодовых последовательностей (криптографических ключей), а также оценка их качества внутри доверенной вычислительной среды [1, 2].

Пользователь не способен точно воспроизвести даже один и тот же рукописный образ. Каждый из биометрических параметров имеет свою длину значимых разрядов (до старшего значимого разряда), последний младший бит является случайным. В силу этого, нестабильная (неповторяемая, случайная) компонента достаточно легко может быть выделена из биометрических данных обучения его нейросетевого преобразователя биометрия-код.

Эту ситуацию поясняет рис. 1, где отображены два написанных почерком пользователя слова «*Пенза*».

Для того, чтобы усилить шумовые свойства нестабильной части биометрических данных над ними необходимо выполнить хэширование. В случае, если доверенная вычислительная среда построена на базе мощного процессора, то хэширование может быть криптографическим [3].

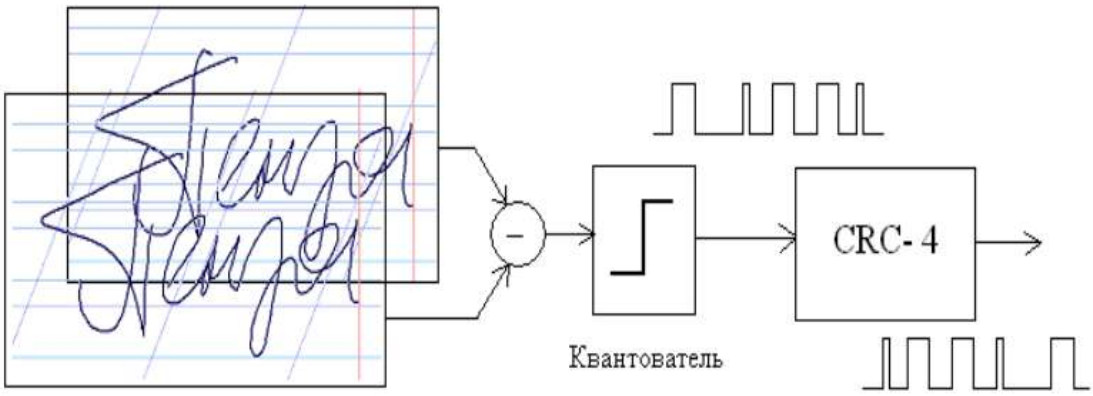


Рис. 1. Выделение не стабильной части биометрических данных, их оцифровывание и хэширование

Если использован процессор с малым потреблением энергии, с малым числом разрядов, с малым объемом оперативной и постоянной памяти (например, RFID идентификационной карты, SIM карты), то приходится использовать ослабленные процедуры защиты информации. В связи с этим на рис. 1 показано использование в качестве простейшего хэширования процедура CRC-4 подсчета контрольной суммы [4].

Таким образом, из нестабильной части биометрических данных пользователя могут быть получены достаточно большие объемы нестабильных, неповторяющихся компонент биометрических данных, хэширование которых (криптографическое или не криптографическое) способно давать «сырой» ключевой материал достаточного объема. Далее для того, чтобы получить ключи достаточно высокого качества необходимо выполнить внутри доверенной вычислительной среды их тестирование [5]. Например, для этой цели могут быть использованы тесты NIST, любо иные популярные тесты.

Большинство известных тестов ориентированы на применение мощных вычислителей, так как обладают экспоненциальной вычислительной сложностью. Естественно, что такие тесты нельзя применять в доверенной вычислительной среде с малопотребляющим процессором.

В связи с этим начали активно развиваться системы тестирования с почти линейной вычислительной сложностью [5, 6, 7, 8], построенные на понижении требований к вычислительным ресурсам за счет перехода от анализа обычных кодов к расстояниям Хэмминга. При таком подходе снимается проблема вычислительной сложности, реализации тестов, однако остро встает вопрос

о том, на сколько была утрачена мощность самих процедур тестирования.

Плохая линейная разделимость зависимых и независимых данных при их анализе в пространстве расстояний Хэмминга

Одной из проблем тестирования больших нейронных сетей с 256 выходами является необходимость применения очень больших тестовых баз. Проблема усугубляется тем, что сбор баз «Чужих» биометрических образов законодательно ограничен практически во всех развитых странах. Без согласия на обработку персональных данных нельзя собирать и хранить «Чужие» биометрические образы.

Выход из этого технологического тупика дает отечественный стандарт [9]. Этот стандарт рекомендует отказаться от статистического анализа большого числа выходных кодовых состояний. В место прямого статистического анализа поля возможных кодовых состояний стандарт рекомендует перейти в пространство расстояний Хэмминга до кода образа «Свой», на который была обучена нейросеть:

$$"h" = \sum_{i=1}^{256} ("c_i") \oplus ("x_i") \qquad (1)$$

где $"c_i"$ – состояние i-го разряда кода «Свой»; $"x_i"$ – состояние i-го разряда, анализируемого кода «Чужой»; $\oplus$ – операция сложения по модулю два бинарных разрядов двоичного числа.

В результате вычисления свертки Хэмминга (1) о суммировании большого числа случайных переменных происходит нормализация задачи и ее экспоненциальной упрощение [10, 11]. На рис. 2 представлены распределения расстояний Хэмминга для шума независимых данных и случайных данных с существенной зависимостью.

Для вычисления математического ожидания и стандартного отклонения вполне достаточно 30 опытов. То есть для достаточно надежного быстрого тестирования нейросети в рамках гипотезы нормального распределения расстояний Хэмминга вполне достаточно малых тестовых выборок.

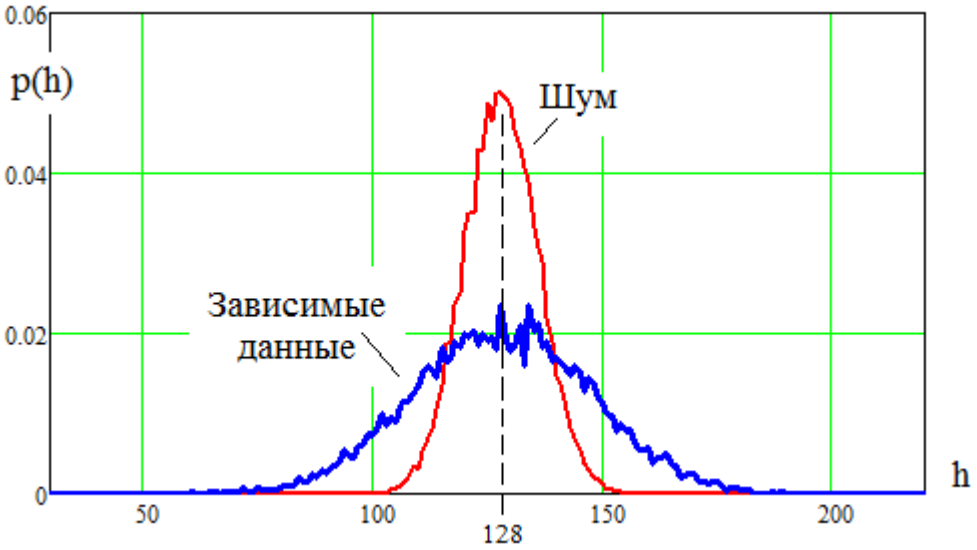


Рис. 2. Отсутствие линейной разделимости шума и зависимых данных в пространстве расстояний Хэмминга

При этом вероятность ошибок второго рода (принятия «Чужого» как пример образа «Свой») оценивается по следующей формуле:

$$P_2 \approx \frac{1}{\sigma(h) \cdot \sqrt{2 \cdot \pi}} \int_0^1 exp\left\{\frac{-(u - E(h))^2}{2 \cdot (\sigma(h))^2}\right\} du \qquad (2)$$

Формально при попадании расстояния Хэмминга в точку «h»≡«0» злоумышленник угадывает код «Свой» подбором. Если мы имеем дело с настоящим «белым» шумом, то угадывание кода «Свой» маловероятно. В этом случае математическое E(h) = 128 бит, а стандартное отклонение должно составлять σ(h) = 8 бит. В теории идеальный «белый» шум должен иметь энтропию 256 бит. Формально многомерная энтропия зависимых кодов «Чужой» может быть оценена следующим образом:

$$H("x_1, x_2, \ldots, x_{256}") \approx -log_2(P_2) \qquad (3)$$

Чем больше корреляционная сцепленность кодов «Чужой», тем больше стандартное отклонение распределений расстояний Хэмминга и тем меньше их энтропия (3). С одной стороны на этом могут быть построены, соответствующие, статистические критерии проверки гипотезы независимости данных [6, 7, 8], а с другой стороны плотность распределения значений расстояний Хэмминга

(рис. 2) находится внутри плотностей распределения значений зависимых данных. Наблюдается плохая линейная разделимость данных «белого» шума и зависимых данных. Это косвенно свидетельствует о относительно низкой мощности системы статистических критериев, построенных на вычислении расстояний Хэмминга по разным модулям [6, 7, 8].

Автокорреляционный статистический критерий проверки гипотезы независимости дискретных выборок

Достаточно случайные данные могут быть получены от любого из программных генераторов. Эти данные в первом приближении могут рассматриваться как эталонные данные «белого» шума. Для сравнения с ними необходимо иметь некоторый эталон зависимых данных, который можно получить, воспользовавшись, скользящим окном шириной в 8 случайных отсчетов:

$$\tilde{x}_j = \frac{1}{8}\sum_{i=0}^{7} x_{j+i} \qquad (4)$$

Если мы вычислим коэффициенты автокорреляции:

$$r(x_j, x_{j+1}) = \frac{E\{(x_j - E(x) \cdot (x_{j+i} - E(x)\}}{\sigma(x) \cdot \sigma(x)} \qquad (5)$$

для эталонного шума и эталонных зависимых данных, то мы получим соотношения, хорошо соответствующие классике. Данные имитационного моделирования для последовательностей в 256 бит представлены на рис. 3.

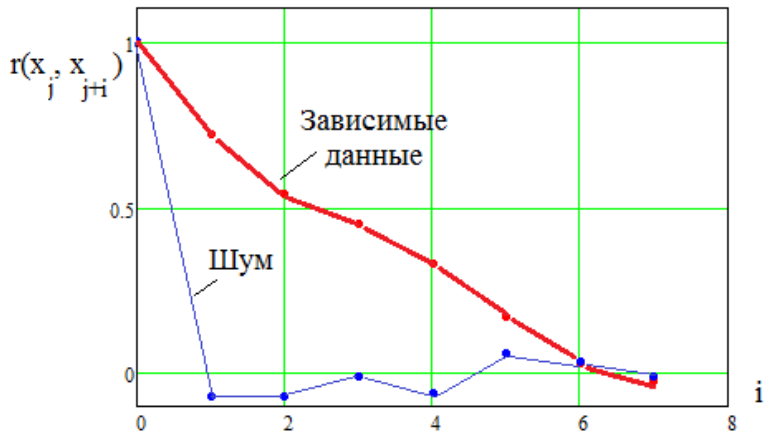


Рис. 3. Хорошее совпадение значений коэффициентов автокорреляции шума (данные программного генератора случайных чисел) и зависимых данных

Заметим, что наклон автокорреляционной функции зависимых данных всегда будет связан с шириной окна сглаживания (4). Чем больше будет ширина окна, тем медленнее будет снижение функции автокорреляции зависимых данных. Автокорреляционная функция шума при любом сдвиге должна давать малые случайные значения, изменяющиеся в близи нулевого состояния.

Так как автокорреляционная функция случайного шума значительно меньше значений автокорреляционной функции зависимых данных, то можно построить модуль-автокорреляционный статистический критерий оценки корреляционной сцепленности длинных кодов:

$$\Sigma|r| = \sum_{i=1}^{7} \left| \frac{E\{(x_j - E(x) \cdot (x_{j+i} - E(x)\}}{\sigma(x) \cdot \sigma(x)} \right| \qquad (6)$$

На рис. 4 приведены результаты численного моделирования эффективности нового статистического критерия (6).

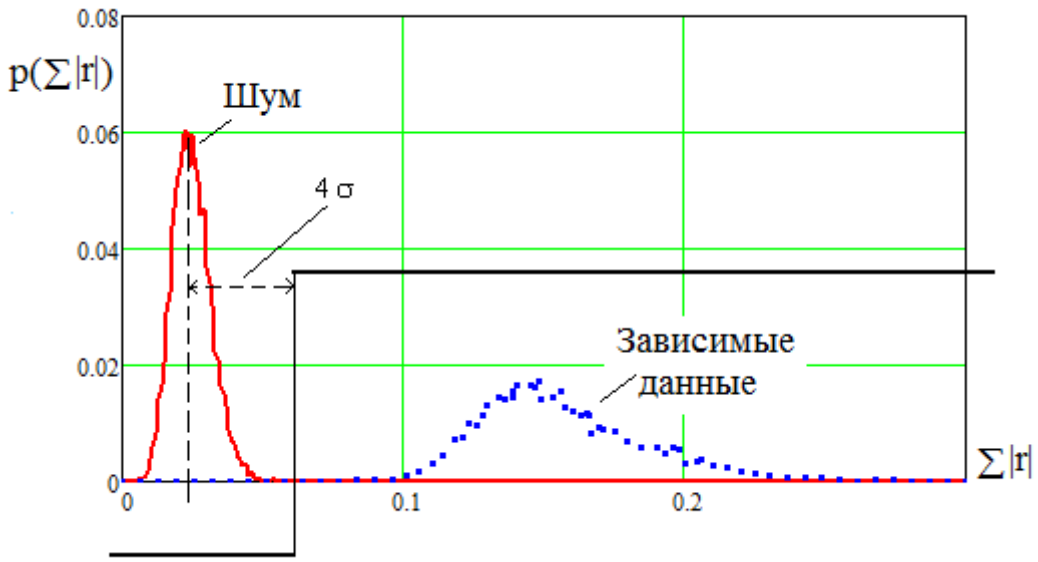


Рис. 4. Высокая линейная разделимость шума и зависимых данных в пространстве автокорреляционных сверток длинных кодов

Из рис. 4 видно, что в пространстве нового статистического критерия распределение «белого» шума и распределение зависимых данных оказались легко разделимы. Это означает значительно повышение мощности нового статистического критерия в сравнении с ранее построенными статистическими критериями пространства множества расстояний Хэмминга.

Если предположить, что квадратичным решающим правилом удастся разделить данные белого шума и зависимые данные (рис. 4) с вероятностями ошибок первого и второго рода $P_1 \approx P_2 \approx P_{EE} \approx 0.2$, то мы получим примерно 1000 кратный выигрыш по мощности для нового статистического критерия $P_1 \approx P_2 \approx P_{EE} \approx 0.0002$. Столь высокий выигрыш обусловлен ситуацией, отображенной на рис. 4, где расстояние от центра распределения «белого» шума до порога принятия решения близко к четырем стандартным отклонениям распределения «белого» шума.

Проведенные исследования показали, что задача вычисления автокорреляционных функций имеет квадратичную сложность и в этом отношении новый класс статистических критериев хуже ранее, исследованных критериев, построенных в пространствах сверток Хэмминга [8]. Тем не менее, высокая потенциальная мощность нового класса статистических критериев дает возможность заменить одним сильным критерием десятки более слабых статистических критериев.

Список литературы

1. Строков А. В., Казанцев Е. И. Программное средство создания действительно случайных криптографических ключей из неоднозначной компоненты биометрических данных динамики рукописного подчерка пользователя // Безопасность информационных технологий : сб. науч. ст. по материалам I Всерос. науч.-техн. конф. (Пенза, 24 апреля 2019 г.). Пенза, 2019. С. 139–144.

2. Юнин А. П., Иванов А. И., Строков А. В., Махсудов С. Р. Нейросетевое обобщение трех стандартных тестов контроля качества «белого шума», полученного хэшированием случайной части биометрических данных // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. (Пенза, 24 июля 2020 г.). Пенза, 2020. С. 31–36.

3. URL: https://ru.wikipedia.org/wiki/Криптографическая_хэш-функция

4. URL: https://ru.wikipedia.org/wiki/Циклический_избыточный_код

5. Григорьев А. Ю. Методы тестирования генераторов случайных и псевдослучайных последовательностей // Ученые записки УлГУ. Сер.: Математика и информационные технологии. 2017. № 1. С. 22–28.

6. Волчихин В. И., Иванов А. И., Юнин А. П., Малыгина Е. А. Многомерный портрет цифровых последовательностей идеального «белого шума» в свертках Хэмминга // Известия высших учебных заведений. Поволжский регион. Технические науки. 2017. № 4. С. 4–13.

7. Юнин А. П., Иванов А. И., Ратников К. А., Кольчугина Е. А. Оценка качества «белого» шума: реализация теста «стаи обезьян» через множество сверток Хэмминга, построенных на разных системах счисления // Известия высших учебных заведений. Поволжский регион. Технические науки. 2018. № 4 (48). С. 54–64.

8. Иванов А. И., Юнин А. П. Эмбрион искусственного интеллекта: компактная нейросетевая проверка качества случайных последовательностей, полученных из биометрических данных : препринт. Пенза : Изд-во ПГУ, 2020. 48 с. ISBN 978-5- 907364-80-6

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

10. Иванов А. И., Кубасов И. А., Самокутяев А. М. Тестирование больших нейронных сетей на малых выборках // Надежность и качество сложных систем. 2020. № 3 (31). С. 72–79.

11. Иванов А. И. Искусственный интеллект высокого доверия Ускорение вычислений и экономия памяти при тестировании больших сетей искусственных нейронов на малых выборках // Системы безопасности. 2020. № 5. С. 60–62.

УДК 004.056

ТЕСТИРОВАНИЕ НЕЙРОСЕТЕВОГО КОРРЕКТОРА ОШИБОК ВЫЧИСЛЕНИЯ МАТЕМАТИЧЕСКИХ ОЖИДАНИЙ НА МАЛЫХ ВЫБОРКАХ С НОРМАЛЬНЫМ ЗАКОНОМ РАСПРЕДЕЛЕНИЯ ЗНАЧЕНИЙ

А. И. Иванов[1], Е. А. Малыгина[2], А. А. Строителева[3], Н. А. Папуша[4], М. А. Митрохин[5]

[1] Пензенский научно-исследовательский электротехнический институт, г. Пенза
[2] Московский государственный технологический университет – МИРЭА, г. Москва
[3,4,5] Пензенский государственный университет, г. Пенза

Аннотация. Приведены результаты тестирования нейросетевого предсказателя ошибок вычисления математического ожидания, возникающих из-за малого объема выборок в 16 опытов. Нейросетевой предсказатель ошибок с 22 выходами рассматривается как черный ящик, заранее обученный распознавать наиболее вероятные значения ошибок. Обучение нейропредсказателя могло быть выполнено любым из известных методов. Показано, что тестируемый программный продукт позволяет снизить стандартное отклонение математического ожидания на 47 %. Это эквивалентно увеличению объема исследуемой выборки с 16 опытов до 35 опытов при обычном способе вычисления математического ожидания. Дана оценка коммерческой ценности нейросетевых предсказаний будущего продукта.

Ключевые слова: нейросетевое предсказание ошибок, математическое ожидание, сортировка данных обучения, муаровые таблицы адресации при сортировке

TESTING A NEURAL NETWORK ERROR CORRECTOR FOR CALCULATING MATHEMATICAL EXPECTATIONS ON SMALL SAMPLES WITH NORMAL DISTRIBUTION LAW OF VALUES

A. I. Ivanov[1], E. A. Malygina[2], A. A. Stroiteleva[3], N. A. Papusha[4], M. A. Mitrokhin[5]

[1] Penza Scientific Research Electrotechnical Institute, Penza
[2] Moscow State Technological University – MIREA, Moscow
[3,4,5] Penza State University, Penza

Abstract. The results of testing a neural network predictor of errors in calculating the mathematical expectation that arise due to a small sample size of 16 experiments

are presented. A neural network error predictor with 22 outputs is treated as a black box, pre-trained to recognize the most likely error values. The neuropredictor could be trained by any of the known methods. It is shown that the tested software product can reduce the standard deviation of the mathematical expectation by 47 %. This is equivalent to increasing the volume of the study sample from 16 experiments to 35 experiments using the usual method of calculating the mathematical expectation. The commercial value of neural network predictions of a future product is assessed.

Keywords: neural network error prediction, mathematical expectation, sorting training data, moire addressing tables during sorting

Введение

Достоверные статистические оценки обычно удается получить на достаточно больших выборках в 200 и более опытов [1, 2]. Во многих исследованиях (медицины, ботаники, биологии, биометрии, экономики) получить столь большие выборки затруднительно. Обычно исследователям легко доступны малые выборки в 16 и более опытов. Выход из создавшегося положения состоит в использовании нейросетевых аналогов различных статистических критериев [3, 4, 5].

Кроме использования статистических критериев часто возникает необходимость в вычислении математического ожидания на малых выборках. В случае, когда выборки исходных данных велики вычисление математических ожиданий осуществляется достаточно точно. Однако на малых выборках положение меняется. Так для малой выборки в 16 опытов псевдослучайный генератор нормальных чисел с программно-заданным нулевым математическим ожиданием дает интервал оценок от –0.8 до +0.8. Распределение значений ошибок приведено на рис. 1.

Очевидно, что повышение числа опытов в выборке должно приводить к снижению интервала неопределенности оценок математического ожидания. Так при увеличении выборки до 35 опытов интервал неопределенности оценок снижается на 47 %. В этом случае интервал неопределенности составит существенно меньшую величину от –0.53 до +0.53.

Так как на практике увеличить выборку реальных данных не всегда возможно, возможен иной путь решения проблемы. Формально мы можем создать большую сеть искусственных нейронной и обучить ее предсказывать ошибку вычисления математического ожидания – ΔE для той или иной выборки. Пример структуры такой сети с 22 выходами приведен на рис. 2.

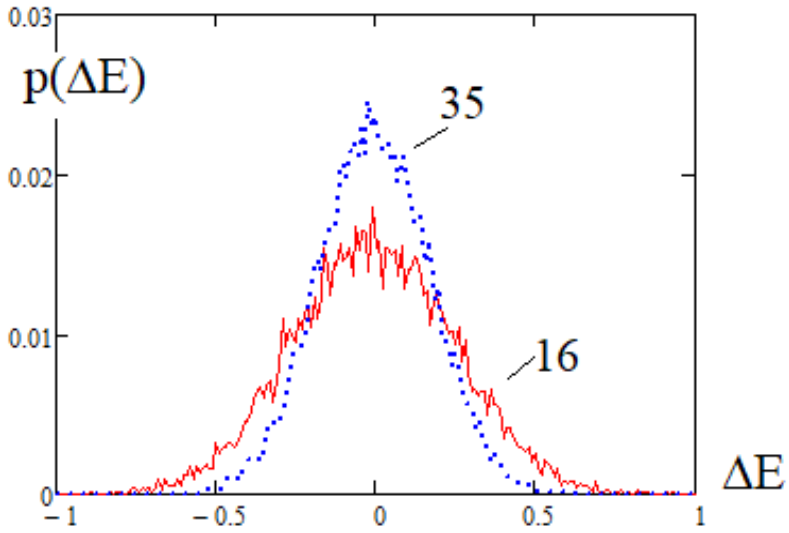


Рис. 1. Распределение ошибок, возникающих
из-за использования малой выборки в 16 опытов
и выборки увеличенной до 35 опытов

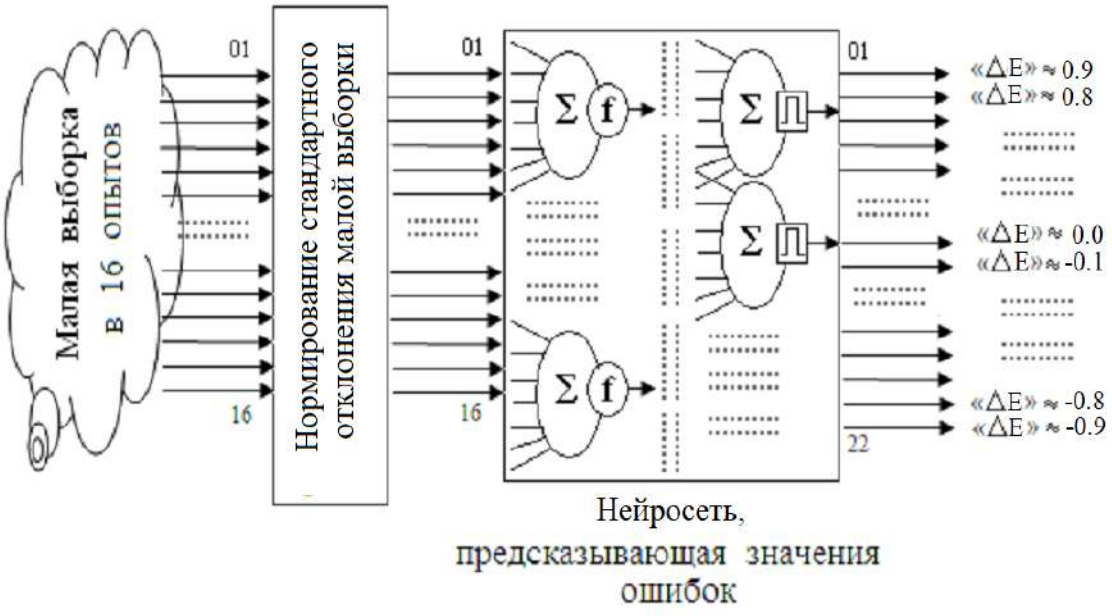


Рис. 2. Нейросетевой предсказатель ошибок вычисления
математического ожидания на малых выборках в 16 опытов

Обладая подобной сетью предсказания ошибок, мы можем использовать ее отклики для коррекции результата вычислений, полученного по классической формуле:

$$E = \frac{1}{16}\sum_{i=1}^{16} x_i \qquad (1)$$

Одним из основных вопросов создания нейросети, способной предсказывать значения ошибок является ее обучение на достоверных данных. Чем больше объем обучающей выборки, тем надежнее должна нейросеть предсказывать ошибки.

В случае, если мы проверяем нейросетью гипотезу нулевого математического ожидания, то любой объем обучающей выборки мы можем получить от генератора псевдослучайных данных с программно заданным значением E = 0.0 и единичным стандартным отклонением σ = 1.0. Соответственно нейросеть сосед с права должна обучаться на данных, полученных от программного генератора с математическим ожиданием E = 0.10.

На текущий момент один из нейросетевых предсказателей ошибок вычисления математических ожиданий создан и, соответственно, может быть выполнено его тестирование. Результаты тестирования приведены на рис. 3.

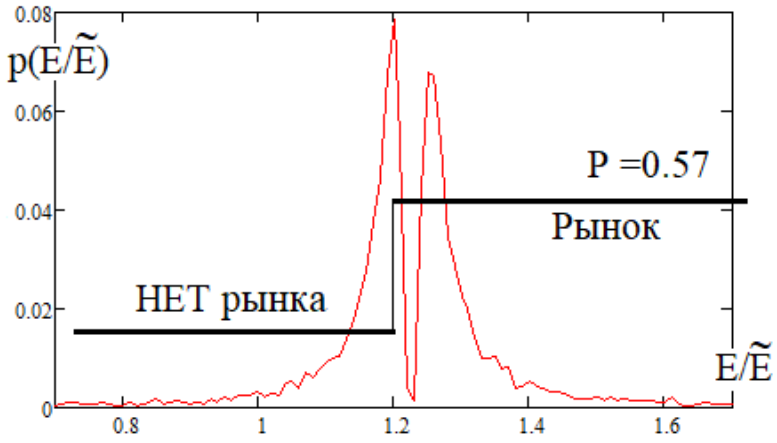


Рис. 3. Доли предполагаемого рынка услуг повышения достоверности оценок математических ожиданий

Из рис. 3 видно, что распределение отношения предсказанных нейросетью значений математических ожиданий и вычисленных по классической формуле (1) имеет два пика. При этом с вероятностью 0.074 нейросетевой предсказатель будет давать

результат хуже, чем классическая формула (1). Очевидно, что этими предсказаниями нельзя пользоваться, гипотетический НейроКалькулятор должен выдавать сообщение «РЕЗУЛЬТАТ ПРЕДСКАЗАНИЯ ХУЖЕ КЛАССИКИ». Так как нейросетевые корректоры являются новым продуктом, они нуждаются в продвижении на рынок. В связи с этим результаты в интервале от 1 до 1.2, видимо, должны предоставляться пользователям бесплатно. Таким образом, коммерчески значимая составляющая (с которой может взиматься плата за услугу) должна появляться с вероятностью 0.57.

Список литературы

1. Кобзарь А. И. Прикладная математическая статистика. Для инженеров и научных работников. М. : ФИЗМАТЛИТ, 2006. 816 с.
2. Дерффель К. Статистика в аналитической химии. М. : Мир, 1994. 258 с.
3. 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
4. Иванов А. И. Нейросетевой многокритериальный статистический анализ малых выборок : справочник. Пенза : Изд-во ПГУ, 2022. 160 с. ISBN 978-5-907600-83-6
5. Иванов А. И. Нейросетевой многокритериальный статистический анализ малых выборок. Проверка гипотезы независимости : справочник. Пенза : Изд-во ПГУ, 2022. 218 с. ISBN 978-5-907666-49-8

УДК 004.056

ОБЗОР МЕТОДОВ ЗАЩИТЫ ДАННЫХ БИОМЕТРИЧЕСКИХ ШАБЛОНОВ

И. Е. Панфилова[1], Д. П. Иниватов[2]

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

Аннотация. Повышенные требования к безопасности для приложений аутентификации привели к экспоненциальному росту систем на основе биометрии. В таких системах сервер хранит биометрическую информацию пользователей, называемую шаблонами. Однако раскрытие биометрических шаблонов в приложениях подобного рода серьезно угрожает конфиденциальности пользователя, так как идентификатором в данном случае выступают его персональные данные. Для решения указанной проблемы биометрические шаблоны подвергаются процедуре обработки специальными алгоритмами, называемыми методами защиты биометрических шаблонов. В представленной работе рассмотрены классификация и особенности функционирования основных подходов к защите биометрических шаблонов.

Ключевые слова: биометрия, защита биометрических шаблонов, гомоморфное шифрование, нечеткие экстракторы, нейросетевой преобразователь биометрия-код, глубокие нейронные сети, отменяемая биометрия

Финансирование: исследование выполнено при финансовой поддержке Минцифры России (грант ИБ), проект № 40469-15/2022-к

OVERVIEW OF BIOMETRIC TEMPLATE DATA PROTECTION METHODS

I. E. Panfilova[1], D. P. Inivatov[2]

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

Abstract. Increased security requirements for authentication applications have led to the exponential growth of biometrics-based systems. In such systems, the server stores users' biometric information, called templates. However, the disclosure of biometric templates in applications of this kind seriously threatens the user's privacy, since the identifier in this case is his personal data. To solve this problem, biometric templates are processed using special algorithms called biometric template protection methods. The presented work examines the classification and operating features of the main approaches to protecting biometric templates.

Keywords: biometrics, biometric template protection, homomorphic encryption, fuzzy extractors, neural network biometric-to-code converter, deep neural networks, cancelable biometrics

Financing: the research was carried out with financial support from the Ministry of Digital Development of Russia (grant IB), project No. 40469-15/2022-k

Введение

Системы аутентификации/идентификации на основе биометрических параметров — это быстро развивающаяся область, в основе которой лежат методы автоматического распознавания людей с помощью их анатомических и/или поведенческих характеристик. Наиболее часто используемыми биометрическими модальностями в наши дни являются отпечатки пальцев, лицо, радужная оболочка, отпечаток ладони, геометрия руки, вена руки, вена пальца и голос. Однако стоит отметить, что рост популярности биометрических систем в самых различных областях человеческой деятельности, в том числе, приводит к появлению новых, повышенных требований к безопасности таких приложений и защите персональных данных, которыми они оперируют. В связи с необходимостью решения подобного рода задач, как в практической, так и исследовательской сферах все больше внимания уделяется методам защиты биометрических шаблонов (ЗБШ) от компрометации.

Отметим, что на сегодняшний день не существует общепринятой классификации методов защиты биометрических шаблонов. Так, например, в работе [1] за основу классификации берется наличие или отсутствие в основе системы распознавания нейронных сетей. В зависимости от того, какую роль играют нейронные сети в структуре системы распознавания лиц, авторы статьи классифицируют методы защиты биометрических шаблонов на не использующие и использующие НС. В первом случае нейронная сеть чаще всего служит исключительно для извлечения признаков из биометрических образов и не является непосредственной частью самого метода защиты биометрического шаблона. При этом извлечение признаков может происходить как до применения методов ЗБШ (тогда осуществляется защита информативных признаков, извлечённых из образа), так и после осуществления защиты входного биометрического образа (тогда признаки извлекаются из уже защищенного образа).

Несмотря на актуальность и удобство приведённой выше классификации методов ЗБШ, данный подход можно упростить, если разделить все существующие методы защиты биометрических шаблонов на четыре основные группы:

1. Методы на основе «связывания» ключа с биометрическим образом.

2. Методы, непосредственно генерирующие бинарный код из биометрического образа.

3. Методы с применением гомоморфного шифрования биометрических шаблонов.

4. Прочие методы.

Полная классификация схем методов ЗБШ представлена на рис. 1.

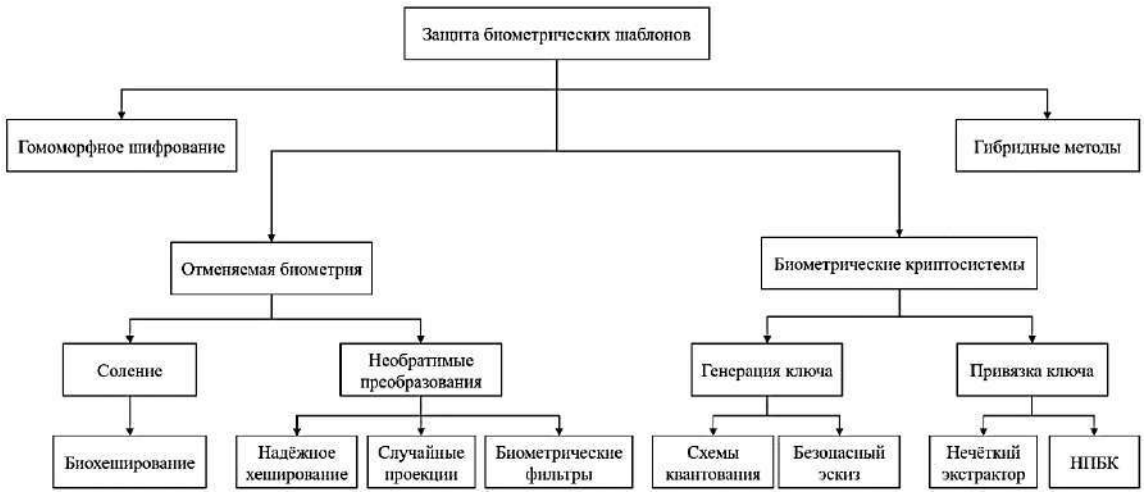


Рис. 1. Иерархическая схема классификации методов ЗБШ

Каждый из указанных методов в той или иной степени позволяет осуществлять защиту биометрических шаблонов путем взаимодействия с системой распознавания в целом или с ее модулями. Работоспособность систем ЗБШ можно оценить по трем категориям [2]: техническая, защитная и эксплуатационная. Технические характеристики можно оценить по точности, пропускной способности, требованиям к хранению. Эксплуатационную производительность можно оценить по независимости от модальности, функциональной совместимости и качеству производительности. К характеристикам защиты относятся необратимость и разнообразие биометрической информации. Согласно стандарту ISO/IEC 24745 «Защита биометрической информации» [3], схема ЗБШ должна удовлетворять четырем требованиям:

– разнообразие: защищенные шаблоны не должны совпадать друг с другом. Это обеспечивает конфиденциальность пользователя;

– возможность отзыва: на основе тех же биометрических данных метод защиты должен иметь возможность отозвать скомпрометированный шаблон;

– необратимость: оригинальные биометрические данные нельзя получить с помощью защищенного (или преобразованного)

шаблона. Вычислительные затраты на получение исходных биометрических данных должны быть невыполнимыми, в то время как создание защищенного шаблона должно быть простым;

– производительность: разработанный метод ЗБШ не должен снижать точность системы распознавания.

Несмотря на неизменность указанных требований, стоит отметить, что их абсолютное выполнение маловероятно в реальных условиях проектирования и функционирования биометрических систем с возможностью защиты биометрического шаблона. Так, например, точность распознавания большинства биометрических систем, как с защитой шаблона, так и без нее, проверяется в «идеальных» условиях, а не в реальных ситуациях, когда распознаваемые образы имеют крайне плохое качество [4]. Потеря значимой информации в исходном образе приводит к низкоэффективной работе лежащих в основе ЗБШ метрик сравнения эталонных шаблонов с полученными в реальном времени, что влечет за собой значительное снижение точности всей системы распознавания в целом.

Методы на основе «связывания» ключа с биометрическим образом

В криптосистеме с «привязкой ключей» выбранный пользователем ключ привязывается к биометрическому шаблону для получения вспомогательных данных. Комбинация ключа и связанного биометрического шаблона хранится в виде защищенного шаблона, который часто называют «вспомогательными данными». Путем подходящей попытки декодирования ключи извлекаются из вспомогательных данных [5].

Обратимый характер алгоритмов шифрования делает их менее привлекательными для защиты конфиденциальных биометрических шаблонов по сравнению с необратимыми операциями, такими как однонаправленные криптографические хеш-функции. Криптографическая хеш-функция создает предсказуемый результат фиксированного размера, называемый «хэшем», из каждого экземпляра одних и тех же входных данных (например, пароля), так что математически невозможно восстановить исходные входные данные из хэша. Несмотря на привлекательность этого свойства «односторонности», применять криптографические хэш-функции к биометрическим шаблонам оказывается достаточно сложно из-за внутриклассовой изменчивости, присущей биометрическим образам. Действительно, хеш-функции специально

разработаны для «преувеличения» небольших различий во входных данных, так что даже их небольшое изменение приведет к совершенно другому выходному хэшу. Учитывая такую чувствительность криптографических хеш-функций к небольшим изменениям во входных данных, методы ЗБШ, использующие криптографическое хеширование, часто избегают создания хэшей из самих биометрических шаблонов. Вместо этого они применяют алгоритм хеширования к случайно сгенерированным внешним кодовым словам, которые связаны с шаблонами какой-либо математической функцией или отображением. Хорошо известным примером такого метода является схема *нечеткого обязательства*, в соответствии с которой случайное, специфичное для пользователя кодовое слово «связывается» с биометрическим шаблоном пользователя. Другим известным типом «привязки» метода ЗБШ является схема *нечеткого контейнера*, где случайно выбранное, специфическое для пользователя кодовое слово служит набором коэффициентов, определяющих секретный полином, а элементы биометрического шаблона пользователя являются входными данными полинома. Указанные подходы могут быть классифицированы как «нечёткие экстракторы», которые обеспечивают надёжную защиту биометрических данных путём создания хэшей, не раскрывающих информацию об исходных шаблонах. Однако, несмотря на их преимущества, нечёткие экстракторы также имеют свои недостатки, которые могут подвергнуть их компрометации. Исследования по этой теме [4] выявили уязвимости нечётких экстракторов, которые могут позволить злоумышленникам восстановить и подделать биометрические шаблоны, что ставит под вопрос эффективность данного подхода в обеспечении безопасности биометрических данных.

Еще одним современным подходом к защите биометрических шаблонов, получившим широкое распространение в отечественном научном и практическом поле, являются преобразователи биометрия-код (ПБК) [6]. Нейросетевая реализация ПБК – «нейросетевой преобразователь биометрия-код» (НПБК), представляет собой однослойную или двуслойную нейронную сеть, способную преобразовывать вектор нечетких, неоднозначных биометрических параметров «Свой» в четкий однозначный код ключа (пароля). Преобразователь откликается случайным выходным кодом на воздействие случайного входного вектора, не принадлежащего множеству образов «Свой». Основные аспекты функционирования НПБК отражены в серии отечественных

стандартов ГОСТ Р 52633 [7]. Особенностью подобного рода реализаций защиты биометрического шаблона является возможность формирования длинных ключей, устойчивых к атакам подбора, а также автоматическое обучение сети нейронов (обучение без участия человека). Такие показатели во многом стали возможны за счет использования специфических реализаций как самой нейронной сети («широкие сети»), так и отдельных нейронов (квадратичные [8], корреляционные [9] и др.). Кроме того, все параметры обученных нейронов (связи и веса) НПБК, называемые нейросетевым контейнером, также оказываются защищены, в том числе от атак извлечения знаний. Для защиты таблиц нейросетевых функционалов применяется механизм защищенного нейросетевого контейнера (ЗНК), отраженный в [10]. В общем виде защита биометрического шаблона в соответствии с указанной спецификацией достигается путем применения использования обратимого или необратимого преобразования.

Отметим, что с целью связывания внешнего кода с биометрическим образом, также могут использоваться глубокие нейронные сети (ГНС). Вариативность их применения в такой задаче во многом сводится к присвоению случайных двоичных кодов максимальной энтропии каждому пользователю системы распознавания лиц, а затем к обучению нейронной сети (сверточной, глубокой, рекуррентной) сопоставлению изображения лица каждого пользователя с соответствующим кодом. Обученная НС должна иметь возможность работать с естественными вариациями биометрических образов (лиц), полученных от одного и того же пользователя, создавая один и тот же код во время регистрации и каждой попытки аутентификации. Двоичные коды, полученные во время регистрации и аутентификации, криптографически хешируются, а сравнение основано на точном совпадении между двумя хэшами. Усовершенствования такого подхода ЗБШ в основном сосредоточены на добавлении дополнительной, специфической для пользователя информации, чтобы усложнить сопоставление между признаками биометрического образа пользователя и его предопределенным (предварительным) кодом. Однако одним из главных недостатков такого подхода является невозможность автоматического обучения указанных ГНС, так как сам процесс обучения, как правило, основывается на методе градиентного спуска, плохо поддающемуся автоматизации. Более того, итерационные методы обучения сетей обладают повышенной склонностью

к переобучению, что также делает их менее надежными для применения в реальных условиях.

Основное преимущество методов рассмотренных методов ЗБШ на основе прямого сопоставления внешнего кода и биометрического образа пользователя путем обучения НС, заключается в том, что результирующие защищенные шаблоны (т.е. криптографические хэши предварительно определенных кодов) принципиально не связаны с исходными (незащищенными) шаблонами лиц. Это означает, что защищенные шаблоны не должны раскрывать информацию о шаблонах лиц, которым они были назначены. Следовательно, если злоумышленник получит доступ к защищенным шаблонам, хранящимся в базе данных системы распознавания лиц, он не сможет восстановить соответствующие изображения или черты лица. Конечно, это основано на предположении, что доступ к параметрам обученной НС не позволит злоумышленнику обнаружить какую-либо связь между шаблоном лица и его хешем. Однако такое предположение может не подтвердиться на практике, особенно если рассматривать наихудший сценарий полностью информированного злоумышленника. Основная проблема с методами ЗБШ, которые полагаются на заранее определенные выходные данные, заключается в том, что НС необходимо будет переобучать (полностью или частично) каждый раз, когда новый пользователь хочет зарегистрироваться в системе распознавания лиц или когда скомпрометированный пользователь должен быть повторно зарегистрирован с новым защищенным шаблоном.

Методы, непосредственно генерирующие бинарный код из биометрического образа

Один из принципов работы таких систем является квантование биометрических данных. Для получения стабильных ключей используются вспомогательные данные. Эти схемы берут векторы признаков нескольких биометрических образцов и получают интервалы элементов признаков. Интервалы кодируются, а затем сохраняются в виде вспомогательных данных. Во время аутентификации биометрические признаки рассчитываются и сопоставляются с определенными интервалами. В результате получается ключ. Так, например, в работе [11] разработали систему «генерации ключей» из отпечатков пальцев, используя нечеткие экстракторы. Однако отметим, что надежное извлечение криптографического ключа из биометрического образа по-прежнему представляет собой

сложную проблему из-за высокой вариативности образов одного человека и ограниченного способа сопоставления полученных образов.

По аналогии с алгоритмом связывания внешнего кода с биометрическим образом, для генерации ключей из биометрических образов также можно использовать нейронную сеть, которая обучается собственному представлению защищенного шаблона лица. Одним из самых ранних примеров такого метода является [12]. Этот метод предлагает глубокую нейронную сеть, состоящую из двух компонентов: компонента глубокого хеширования (ГХ) и компонента декодера нейронной сети. Компонент ГХ обучается генерировать промежуточный двоичный код по входному изображению лица, после чего декодер учится исправлять ошибки в этом двоичном коде. Затем исправленный код криптографически хэшируется для создания защищенного шаблона. Однако, поскольку нейронная сеть обучена изучать один и тот же код для каждого изображения лица от одного и того же пользователя, отзыв скомпрометированных защищенных шаблонов представляется невозможным.

Несмотря на то, что данное направление получило в литературе широкое освещение и, как правило, выделяется в отдельную группу методов, подход, который лежит в основе вызывает много вопросов. Прежде всего, не понятно, каким образом исследователи добиваются необходимой длины и энтропии бинарного кода. При генерации криптографических ключей и паролей применяются определенные практики и стандарты, которые гарантируют уникальность и высокую энтропию генерируемых ключей. Есть обоснованные сомнения в том, что методы непосредственной генерации бинарного кода из биометрических данных могут обеспечить необходимый уровень энтропии и уникальности (и насколько корректно называть генерируемый в этом случае бинарный код криптографическим паролем или ключом). Тем не менее, в обзорной работе нельзя обойти вниманием данную категорию методов.

Методы с применением гомоморфного шифрования биометрических шаблонов

Одним из наиболее часто изучаемых алгоритмов защиты биометрических шаблонов, является гомоморфное шифрование (ГШ), которое было применено к признакам, извлеченным из изображений лиц с использованием предварительно обученных моделей глубокой нейронной сети [13]. В отличие от традиционных методов шифрования, ГШ позволяет выполнять операции

с зашифрованными данными без необходимости их предварительной расшифровки. Несмотря на очевидные преимущества, вычислительная сложность алгоритма не делает его широко применимым в задачах защиты биометрических шаблонов. Кроме того, во многом алгоритм подвержен так называемой проблеме накопления ошибок, так как после совершения множества операций с зашифрованными данными результат может перестать соответствовать результату этих операций с незашифрованными данными. Текущие исследования в этой области, как правило, сосредоточены на поиске баланса между ускорением операций ГШ и одновременной минимизацией потерь в результирующей точности распознавания.

Прочие методы

Еще одним популярным способом защиты биометрических шаблонов является отменяемая биометрия. Отменяемая биометрия – это ряд специальных решений, которые включают преобразование исходных биометрических данных в новую форму, которая не распознается злоумышленниками, но при этом сохраняет свою дискриминационную способность для целей распознавания. Концепция отменяемой биометрии была введена для того, чтобы биометрический шаблон можно было отменить и отозвать, как пароль, а также сделать его уникальным для каждого приложения. Отменяемая биометрия требует хранения «искаженной» версии биометрического шаблона, что обеспечивает высокий уровень конфиденциальности, позволяя связать несколько шаблонов с одними и теми же биометрическими данными.

Так, видами отменяемой биометрии могут выступать соление и необратимые преобразования. Соление – это подход ЗБШ, в котором биометрические характеристики преобразуются с помощью обратимой функции. В этих системах ключ (вспомогательные данные) должен храниться в безопасном месте или вызываться пользователем для аутентификации, поскольку используемое преобразование является обратимым. Скомпрометированные шаблоны легко отозвать, изменив пользовательские ключи. В случае использования таких ключей их необходимо предъявить при аутентификации [14].

Ещё одним методом ЗБШ являются необратимые преобразования. В отличие от методов соления, они применяют алгоритмы,

не дающие возможность осуществить обратную операцию. Параметры функции преобразования, называемые ключом, должны быть созданы во время аутентификации. Исходные биометрические данные не могут быть восстановлены, что повышает безопасность [15]. Например, потеря связана с артефактом, который трудно совместить с биометрическими шаблонами, такими как отпечатки пальцев. Чтобы уменьшить эту трудность, позже были предложены методы без выравнивания для отпечатков пальцев. Кроме того, необратимые преобразования обеспечивают лучшую возможность отзыва и разнообразие по сравнению с подходом с добавлением соли [16]. При разработке необратимых преобразований следует приложить усилия, чтобы сохранить компромисс между различимостью и необратимостью.

Отменяемая биометрия предлагает решение для сохранения конфиденциальности пользователя, поскольку истинная биометрия пользователя никогда не раскрывается в процессе аутентификации. Это гарантирует, что защита шаблона достигается на уровне функций с помощью вспомогательных данных или необратимых преобразований. С другой стороны, отменяемая биометрия имеет определенные ограничения, которые необходимо учитывать. Например, при солении шаблон может перестать быть безопасным, если вспомогательные данные скомпрометированы. Для необратимых преобразований необратимость повышает безопасность шаблонов за счет использования процесса преобразования для сброса порядка или положения набора функций. Однако это ослабляет дискриминационную способность (производительность) преобразованных признаков из-за увеличения внутриклассовой вариации биометрических показателей. В этом контексте, если производительность является главной задачей при проектировании биометрической системы, то ожидается, что в системе будет отсутствовать случайность, необходимая для проектирования безопасного и непредсказуемого пространства шаблонов.

Гибридные методы BTP могут быть разработаны путем объединения отменяемой биометрии и криптосистем. Схемы, описанные выше, комбинируются с использованием любых биометрических модальностей. Так, авторы работы [17] представили схему для лицевой биометрии. Этот метод делит данные на две части: «дробную часть» и «целую часть». «Дробная часть» выполняет преобразование, а «целая часть» шифруется. В работе [18] авторы расширили подход на биометрические образы в виде отпечатков пальцев.

Заключение

Безопасность процедуры биометрической аутентификации является критическим вопросом, поскольку существует множество аспектов, связанных с целостностью и общедоступностью систем аутентификации на основе биометрических данных. В связи с этим, защита биометрических шаблонов от различных хакерских атак представляет особый интерес при проектировании и разработке биометрических систем распознавания (в отличие от паролей или токенов, скомпрометированные биометрические шаблоны не подлежат повторному выпуску).

Среди возможных подходов к решению проблемы защиты биометрических шаблонов от компрометации в представленной работе рассмотрены 4 группы методов: методы на основе «связывания» ключа с биометрическим образом; методы, непосредственно генерирующие бинарный код из биометрического образа; методы с применением гомоморфного шифрования биометрических шаблонов; прочие методы. Предложенная классификация охватывает весь круг существующих на сегодня способов реализации защиты биометрических шаблонов. Исследование также отражает ограничения различных методов защиты биометрических шаблонов, используемых в настоящее время. Среди наиболее перспективных методов ЗБШ выделен нейросетевой преобразователь «биометрия-код», позволяющий не только обеспечивать безопасность биометрического шаблона, но и сохранять высокий уровень точности распознавания биометрических образов.

Список литературы

1. Hahn V. K., Marcel S. Biometric template protection for neural-network-based face recognition systems: A survey of methods and evaluation techniques // IEEE Transactions on Information Forensics and Security. 2022. Т. 18. С. 639–666.

2. Simoens K. [et al.]. Criteria towards metrics for benchmarking template protection algorithms // 2012 5th IAPR International Conference on Biometrics (ICB). IEEE. 2012. С. 498–505.

3. Иванов А. И. [и др.]. О многообразии метрик, позволяющих наблюдать реальные статистики распределения биометрических данных «нечетких экстракторов» при их защите наложением гаммы // Вестник УрФО. Безопасность в информационной сфере. 2014. №. 2 (12). С. 16–23.

4. Sarkar A., Singh B. K. A review on performance, security and various biometric template protection schemes for biometric authentication systems // Multimedia Tools and Applications. 2020.

5. Uludag U. [et al.]. Biometric cryptosystems: issues and challenges // Proceedings of the IEEE. 2004. Т. 92, no 6. С. 948–960.

6. Иванов А. И., Ложников П. С., Сулавко А. Е. Оценка надежности верификации автографа на основе искусственных нейронных сетей, сетей многомерных функционалов Байеса и сетей квадратичных форм // Компьютерная оптика. 2017. Т. 41, № 5. С. 765–774.

7. ГОСТ Р 52633–2006. Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации.

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

9. Иванов А. И., Сулавко А. Е. Использование сетей корреляционных нейронов с многоуровневым квантованием: защита от извлечения знаний из параметров решающего правила : препринт. Пенза : Изд-во ПГУ, 2020.

10. ТС 26.2.002–2020. Защита нейросетевых контейнеров с использованием криптографических алгоритмов.

11. Liu E. [et al.]. Minutiae and modified biocode fusion for fingerprint-based key generation // Journal of Network and Computer Applications. 2010. Т. 33, no 3. С. 221–235.

12. Talreja V., Valenti M. C., Nasrabadi N. M. Zero-shot deep hashing and neural network-based error correction for face template protection // 2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS). IEEE. 2019. С. 1–10.

13. Terhörst P. [et al.]. Beyond identity: What information is stored in biometric face templates? // 2020 IEEE international joint conference on biometrics (IJCB). IEEE. 2020. С. 1–10.

14. Teoh A. B. J., Kuan Y. W., Lee S. Cancellable biometrics and annotations on biohash // Pattern recognition. 2008. Т. 41, no 6. С. 2034–2044.

15. Ratha N. K., Connell J. H., Bolle R. M. Enhancing security and privacy in biometrics-based authentication systems // IBM systems Journal. 2001. Т. 40, № 3. С. 614–634.

16. Jain A. K., Nandakumar K., Nagar A. Biometric template security // EURASIP Journal on advances in signal processing. 2008. С. 1–17.

17. Boult T. Robust distance measures for face-recognition supporting revocable biometric tokens // 7th International Conference on Automatic Face and Gesture Recognition (FGR06). IEEE. 2006. С. 560–566.

18. Boult T. E., Scheirer W. J., Woodworth R. Revocable fingerprint biotokens: Accuracy and security analysis // 2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE. 2007. С. 1–8.

УДК 004.056

ПРОГРАММНОЕ ФОРМИРОВАНИЕ ОДНОМЕРНЫХ ЭТАЛОННЫХ ДАННЫХ МАЛЫХ ВЫБОРОК С ЗАРАНЕЕ ЗАДАННЫМ ПОКАЗАТЕЛЕМ ХЁРСТА

Д. В. Тарасов

Пензенский государственный университет, г. Пенза

Аннотация. Вычисление показателей Хёрста активно используется при исследовании данных личной биометрии и данных коллективной биометрии. Целью статьи является создание эталонных малых выборок с заранее известным показателем Хёрста для последующего тестирования компактного, автономного программного средства, способного работать без использования вызовов внешних коммерческих продуктов. Эталонные данные синтезируются на основе получения двух независимых малых выборок заданного объема с нормальным распределением. Синтез эталонов выполняется на основе хорошо разработанной теории хаоса для броуновского движения на плоскости с разными значениями показателя Хёрста.

Ключевые слова: показатель Хёрста, малые выборки, личная биометрия

SOFTWARE FORMATION OF ONE-DIMENSIONAL REFERENCE DATA OF SMALL SAMPLES WITH A PRE-SET HURST INDICATOR

D. V. Tarasov

Penza State University, Penza

Abstract. The calculation of Hurst exponents is actively used in the study of personal biometric data and collective biometric data. The purpose of the article is to create reference small samples with a previously known Hurst exponent, for subsequent testing of a compact, stand-alone software tool that can work without using calls to external commercial products. The reference data is synthesized by obtaining two independent small samples of a given size with a normal distribution. The synthesis of standards is carried out on the basis of a well-developed chaos theory for Brownian motion on a plane with different values of the Hurst exponent.

Keywords: Hurst index, small samples, personal biometrics

Введение

Показатель Хёрста является одним из активно используемых параметров при анализе рынка [1, 2, 3] и коллективной биометрии

[4]. К сожалению, эмпирический показатель Хёрста по умолчанию предполагает использование больших выборок из-за того, что он является степенным:

$$\frac{R}{\sigma} = \left\{\frac{N}{2}\right\}^{H} \quad (1)$$

где N – размер выборки; R – размах выборки; σ – стандартное отклонение выборки; H – степенной показатель Хёрста, изменяющийся в интервале от 0.5 до 1.0 для предсказуемых персистентных систем и, изменяющийся в интервале от 0.0 до 0.5 для антиперсистентных систем.

Переходя в логарифмическую форму уравнения (1), получим еще один второй вариант записи показателя Хёрста:

$$H = log(R/\sigma)/log(N/2) \quad (2)$$

Вторая форма удобна для пояснения причин, по которым при анализе данных рынков и данных коллективной биометрии необходимы большие выборки. Если предположить, что логарифм нормированного размаха данных является константой, то ошибка оценки показателя будет оцениваться следующим соотношением:

$$\Delta H \approx \text{const}/log(N/2) \quad (3)$$

То есть ошибки из-за сокращения размеров выборки реальных данных должны расти обратно пропорционально логарифму объема выборки.

Кажется, что оценка значения ошибок (3) крайне пессимистична и обрекает исследователей на накопление больших объемов данных. Возможен обходной путь решения проблемы через создание быстрых алгоритмов оценки показателя Хёрста [5, 6]. Этот путь состоит в отказе от попыток наблюдения нормированного размаха экспериментальных данных (1) с замещением на наблюдение поведения автокорреляционных функционалов [7].

Теоретическое обоснование такой возможности строится на том, что высоко персистентные временные ряды с высоким значением показателя Хёрста H ≈ 1.0 должны давать высокие значения классических автокорреляционных функционалов $r(x(t), x(t-\tau)) \approx 1.0$ при любом сдвиге данных. В противоположном случае, когда показатель Хёрста минимален H ≈ 0.0, значения

автокорреляционных функционалов должно иметь случайный знак при высоком значении его модуля $r(x(t), x(t-\tau)) \approx \pm 1.0$ (предельно высокая антиперсистентность). В случае, когда показатель $H \approx 0.5$, автокорреляционные функционалы продолжают иметь случайный знак при малых значениях их модулей $r(x(t), x(t-\tau)) \approx \pm 0.0$.

Моделирование данных броуновского движения

В работах Петерса и Мандельбродта [2, 3] была показана связь параметров броуновского движения со значениями показателя Хёрста. На рис. 1 дана программная реализация и пример траектории броуновского блуждания с независимыми (не коррелированными) приращениями.

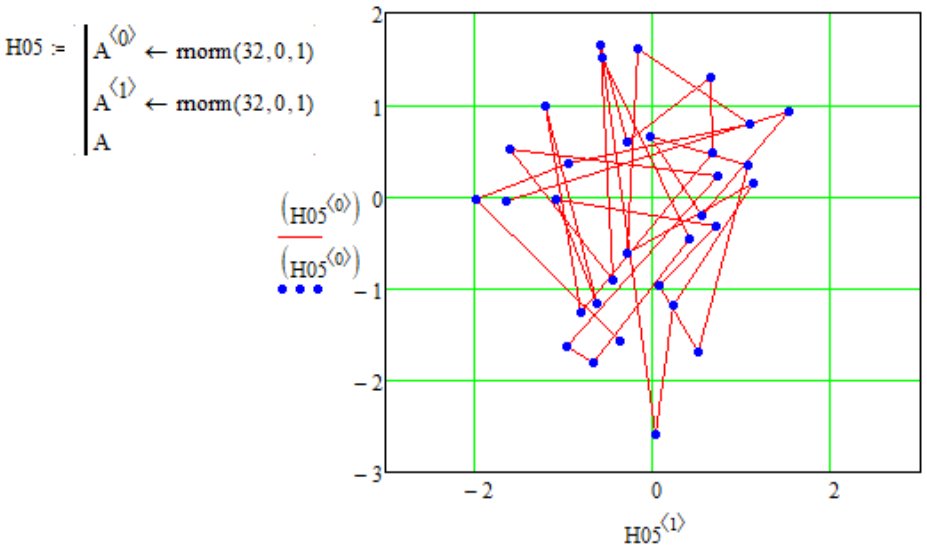


Рис. 1. Результат моделирования броуновского движения H05 (независимые данные)

В первом приближении двухмерное распределение точек рис. 1 должно хорошо описываться окружностью. Если ввести существенную зависимость, анализируемых данных, то ситуация меняется. На рис. 2 представлена программная реализация численного эксперимента и его результаты.

Из данных рис. 2 следует, что распределение точек, анализируемой выборки и броуновские траектории блужданий между точками должны хорошо описываться эллипсом, наклоненным примерно на 45 градусов по малой и большой осям. Задача нейросетевого разделения наклоненного эллипса и окружности вполне посильна для сетей искусственных нейронов [7].

```
H08 := | x ← morm(32,0,1)
       | y ← morm(32,0,1)
       | z ← morm(32,0,1)
       | x08 ← (x + z·2.03)·0.5
       | y08 ← (y + z·2.03)·0.5
       | A^⟨0⟩ ← x08
       | A^⟨1⟩ ← y08
       | A
```

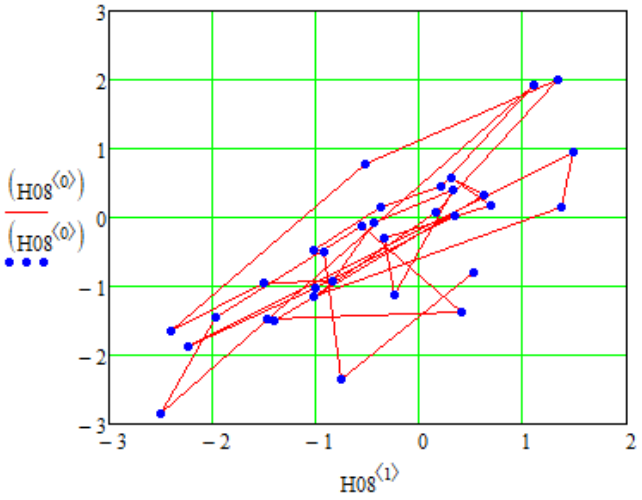


Рис. 2. Результат моделирования броуновского движения H08

Если рассматривать броуновское движение с отрицательно коррелированными данными и показателем Хёрста H02, то распределение точек и траекторий будет описываться эллипсом повернутым на 90 градусам по большой и малым осям по отношению к предыдущему эллипсу H08.

```
H02 := | x ← morm(32,0,1)
       | y ← morm(32,0,1)
       | z ← morm(32,0,1)
       | x02 ← (x + z·2.03)·0.5
       | y02 ← (y − z·2.03)·0.5
       | A^⟨0⟩ ← x02
       | A^⟨1⟩ ← y02
       | A
```

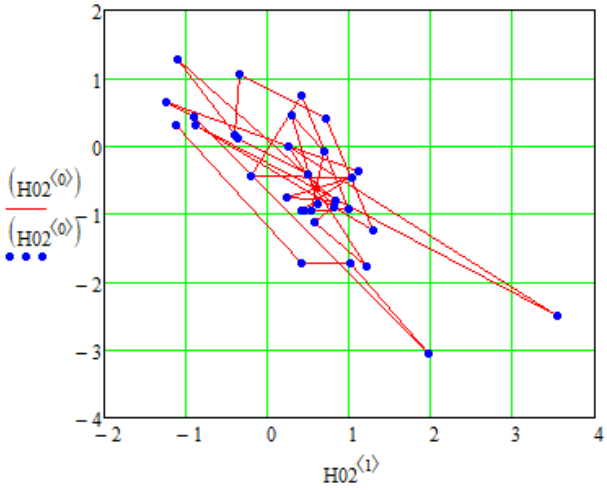


Рис. 3. Результат моделирования броуновского движения H02

В рамках броуновских случайных блужданий задача нейросетевого распознавания значений показателей Хёрста оказывается достаточно простой, что является предпосылкой для ее решения при анализе реальных данных рынка и биометрии.

Программное формирование эталонов малых выборок с заранее заданным показателем Хёрста

Как следует из предыдущего раздела численное моделирование двухмерного броуновского движения является удобным

инструментом, пользующимся высоким доверием к нему научно-технической общественности. Однако идеальные условия хаоса двухмерного броуновского движения не соответствуют ситуации анализа данных, получаемых на рынке. При анализе данных рынка и коллективной биометрии мы имеем доступ к временным рядам, являющихся свертками двухмебрного броуновского хаоса.

Для того, чтобы получить эталонные одномерные выборки необходимо выполнить свертывание двухмерных данных. Эта операция выполнена для низкого уровня показателей Хёрста программно, сама программная реализация приведена в левой части рис. 4.

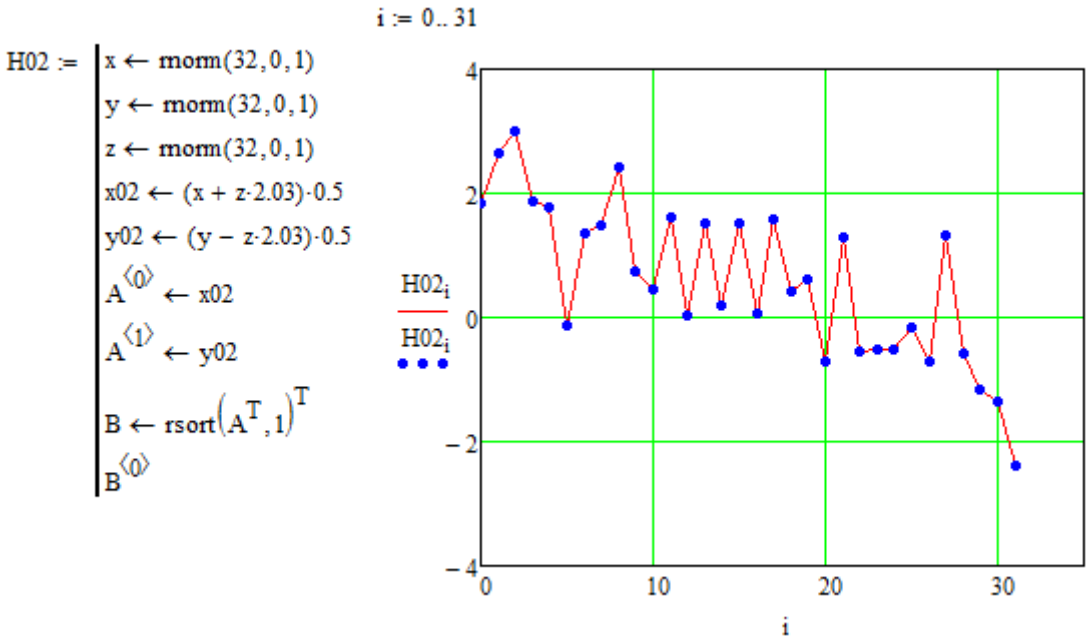


Рис. 4. Пример малой выборки свернутого броуновского движения H02

Как видно из рис. 4 наблюдается тренд падения контролируемых параметров. Пользуясь этим механизмом, мы модем для показателя Херста H02 создать нужное число примеров, опираясь на которые мы можем обучить некоторый нейросетевой классификатор малых выборок в 32 опыта. Сколько подобный классификатор должен иметь слоев искусственных нейронов – это открытый вопрос. Какова доверительная вероятность к подобным нейровычислениям так же является открытым вопросом. Однако существующий положительный опыт обучения нейронных сетей свидетельствует о том, что рост объемов обучающей выборки и рост числа слоев нейронов в сети приводят к росту доверительной

вероятности к принимаем решениям. Так промышленно используемые сегодня сети из сотен слоев искусственных нейронов, обученные распознаванию геометрических особенностей лиц людей позволяют различать образ «Свой» с доверительной вероятностью 0.997.

Предположительно, что для практики распознавания малых выборок данных рынка в 32 опыта доверительной вероятности на уровне 0.95 будет вполне достаточно.

Список литературы

1. Калуш Ю. А., Логинов В. М. Показатель Хёрста и его скрытые свойства // Сибирский журнал индустриальной математики. 2002. Т. 5, вып. 4. С. 29–37

2. Петерс Э. Хаос и порядок на рынках капитала. Новый аналитический взгляд на циклы, цены и изменчивость рынка / пер. с англ. В. И. Гусевой. М. : МИР, 2000. 333 с. ISBN 5-03-003356-4 (рус.), ISBN 0-471-13938-6 (англ.)

3. Мандельброт Б., Хадсон З. Л. (НЕ)послушные рынки. Фрактальная революция в финансах. М. ; СПб. ; Киев : Вильямс, 2006. 408 с. ISBN 5-8459-0922-8, 0-1300-9717-9

4. Иванов А. И. Высокоразмерная коллективная биометрия подсознательного поведения людей на рынке и производстве : препринт. Пенза : Изд-во ПГУ, 2021. 60 с. ISBN 978-5-907456-44-0

5. Иванов А. И., Егорова Ю. Ю. Корреляционный метод быстрой оценки текущего значения показателя Хёрста биометрических данных и данных рынка // Нейрокомпьютеры: разработка, применение. 2012. № 3. С. 26–27.

6. 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. ISBN 978-5-907364-24-0

7. Иванов А. И. Нейросетевой многокритериальный статистический анализ малых выборок. Проверка гипотезы независимости : справочник. Пенза : Изд-во ПГУ, 2022. 218 с. ISBN 978-5-907666-49-8

УДК 004.056

СНИЖЕНИЕ ВЕРОЯТНОСТИ РЕАЛИЗАЦИИ УГРОЗЫ АТАК МАРШАЛКО ПРИ ПЕРЕХОДЕ ОТ ИСПОЛЬЗОВАНИЯ БИНАРНЫХ НЕЙРОНОВ К ТРОИЧНЫМ ИСКУССТВЕННЫМ НЕЙРОНАМ

К. Н. Савинов

Пензенский государственный университет, г. Пенза

Аннотация. Показано, что одним из преимуществ перехода от бинарных искусственных нейронов к троичным искусственным нейронам является снижение вероятности реализации атак Маршалко по сужению поля перебора состояний выходного ключа нейросетевого преобразователя биометрии. Показано, что для соседних троичных нейронов с линейным накоплением данных допустимы одиночные общие связи. Выдвинута гипотеза о том, что для четверичных нейронов должны быть допустимы парные общие связи у соседних нейронов.

Ключевые слова: атака Маршалко, бинарные нейроны, троичные нейроны

REDUCING THE PROBABILITY OF IMPLEMENTATION OF THE THREAT OF MARSHALKO ATTACKS DURING THE TRANSITION FROM THE USE OF BINARY NEURONS TO TRINARY ARTIFICIAL NEURONS

K. N. Savinov

Penza State University, Penza

Abstract. It is shown that one of the advantages of the transition from binary artificial neurons to ternary artificial neurons is the reduction in the likelihood of implementing Marshalko attacks to narrow the field of enumerating the states of the output key of the neural network biometrics converter. It is shown that single common connections are allowed for neighboring ternary neurons with linear data accumulation. It is hypothesized that for quaternary neurons, pairwise common connections among neighboring neurons should be allowed.

Keywords: Marshalko attack, binary neurons, ternary neurons

Введение

Первым в мировой практики стандартом по быстрому автоматическому обучению сетей искусственных нейронов на малых

выборках является ГОСТ Р 52633.5–2011 [1]. На рис. 1 приведена типовая структура применения этого стандарта к обработке биометрических данных пользователя.

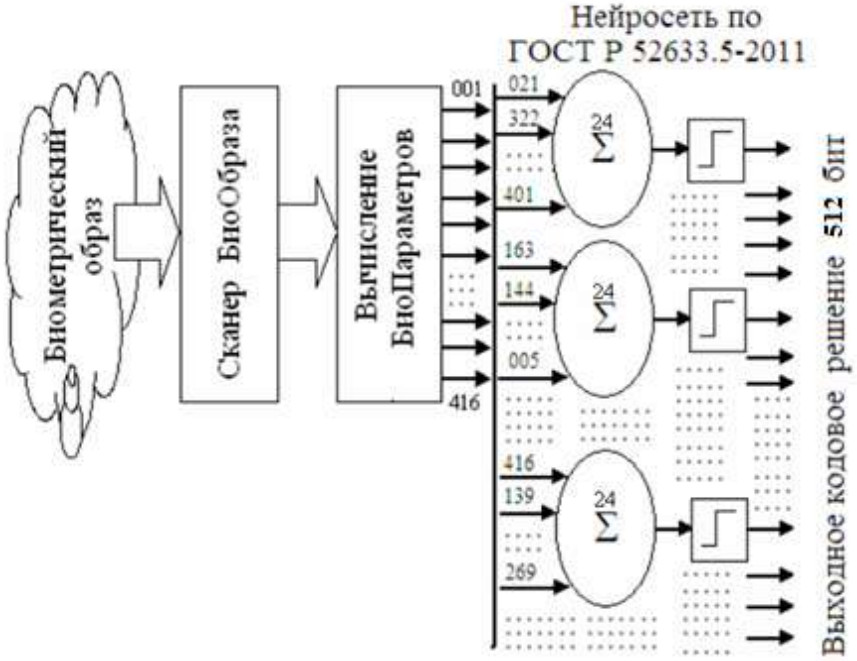


Рис. 1. Типовая структура преобразования данных в соответствии с ГОСТ Р 52633.5–2011 и рядом других стандартов этого же пакета

В случае, когда обученная нейронная сеть должна безопасно храниться (пересылаться по открытому каналу), ее параметры должны быть зашифрованы. Шифрование должно вестись по технической спецификации ТК 26 [2]. Фактически речь идет о создании самораспаковывающегося зашифрованного архива. При этом часть данных для последовательного расшифровывание данных архива получается из входного ключа расшифровывания, а другая часть ключа расшифровывания берется из выходных данных уже расшифрованной части нейросети.

Одной из проблем технической спецификации является то, что она ориентирована на усеченную версию нейросети, обученной по стандарту [1]. Первые нейроны такой сети не должны иметь входных общих связей из-за угрозы применения атак Маршалко [3, 4].

Атака Маршалко

Рассмотрим ситуацию, когда входные параметры пронумерованы монотонно возрастающей последовательностью цифр, что

отображено на рис. 2. По стандарту [1] входные параметры каждого нейрона выбираются случайно. В этом случае возможно попадание в одну группу двух и более нейронов, имеющих общие входные связи, как это показано на рис. 2.

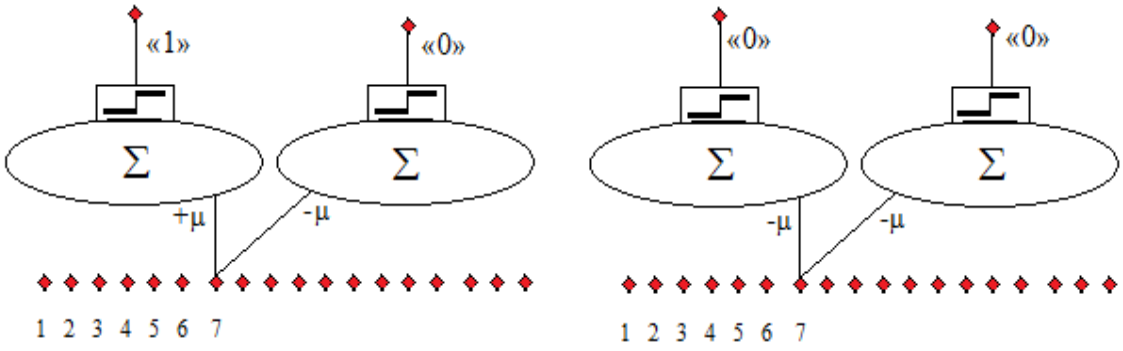


Рис. 2. Сортировка нейронов при реализации атаки Маршалко через обнаружение общих связей в группах нейронов

При этом в левой части рисунка отображена ситуация, когда общие связи двух нейронов имеют весовые коэффициенты разного знака. В этом случае нейроны с одной общей связью должны иметь разные выходные состояния.

В правой части рисунка отображена ситуация, когда два нейрона имеют одну общую связь при одинаковых весовых коэффициентах. Тогда оба нейрона должны иметь одинаковые выходные отклики.

Таким образом, возможно существенное сокращение числа выходных состояний ключа, образующегося на выходах нейросети при воздействии на нее биометрическими данными образа «Свой» за счет группировки нейронов по числу у них входных общих связей. Для противодействия угрозе реализации атак Маршалко необходимо как минимум контролировать наличие в защищаемой криптографией нейросети наличие групп нейронов с 1, 2, 3, … общими связями [5]. При обнаружении общих входных связей у бинарных нейронов, приходится исключать такие нейроны.

Переход от бинарных искусственных нейронов к троичным

Положение меняется, если усложнить искусственные нейроны, заменив их бинарный выходной квантователь на троичный. Эта ситуация отображена на рис. 3.

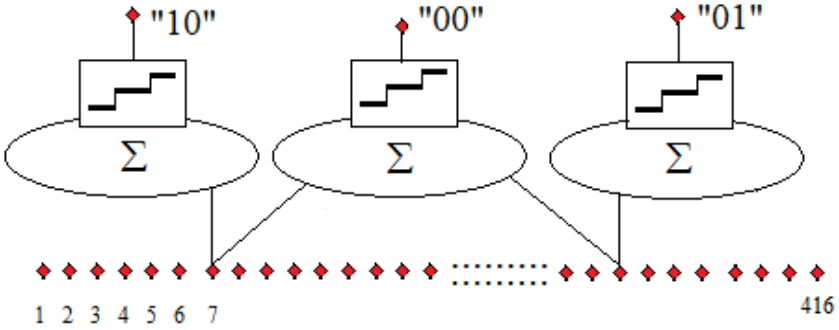


Рис. 3. Пример ситуации, когда нельзя различить (нельзя, верно, группировать) нейроны, имеющие одинаковые связи, но совершенно разные отклики

На рис. 3 отображена ситуация, когда три нейрона, имеющие общие связи с другими нейронами имеют совершенно разные выходные состояния, при одинаковых весовых коэффициентах $\mu = 1$.

Первым в мировой практике был разработан стандарт автоматического обучения квадратично-троичных искусственных нейронов [6]. При этом введение в искусственный нейрон троичного квантователя существенно меняет ситуации.

Следующим стандартом с троичными нейронами, видимо, должен стать стандарт по автоматическому обучению троичных нейронов с линейным накоплением [7]. В этом случае вполне допустимо иметь группы нейронов с одной общей связью. Экспериментально доказано, что для таких нейронов нельзя предсказать их выходные состояния. Численный эксперимент выполнялся на макете нейросети «БиоНейроАтограф III» с 8 и 16 входовыми нейронами.

Предположительно, что два нейроны, имеющие по две общие связи вновь будут подвержены атакам Маршалко, однако это предположение нуждается в подтверждении.

Список литературы

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

2. Техническая спецификация ТС 26.2.002–2020. Системы обработки информации. Криптографическая защита информации. Защита нейросетевых биометрических контейнеров с использованием криптографических алгоритмов. URL: https:// tc26.ru

3. Маршалко Г. Б. Вопросы оценки стойкости нейросетевой системы биометрической аутентификации» : материалы науч.-техн. конференции «РусКрипто-2013» (29 марта, г. Москва). URL: http://www.ruscrypto.ru/netcat_files/File/ruscrypto.2013.051.zip

4. Marshalko G. B. On the security of a neural network-based biometric authentication scheme // Математические вопросы криптографии. 2014. Т. 5 (2). P. 87–98.

5. Иванов А. И., Крохин И. А. Таблица вероятности появления разных стартовых условий для атак Маршалко на нейроны с общими входными связями // Состояние и перспективы развития современной науки по направлению «Техническое зрение и распознавание образов» : сб. ст. III науч.-техн. конф. Анапа, 2021. С. 171–172.

6. Проект стандарта «Искусственный интеллект. Нейросетевые алгоритмы в защищённом исполнении. Автоматическое обучение нейросетевых моделей на малых выборках в задачах классификации».

7. Защищенные приложения искусственного интеллекта: модификация алгоритма автоматического обучения бинарных персептронов по ГОСТ Р 52633.5–2011 под троичные искусственные нейроны : отчет по НИР (промежуточный). Пенза, 2023. 43 с.

УДК 004.056

ОЦЕНКА АППАРАТНЫХ И ЭНЕРГЕТИЧЕСКИХ НАКЛАДНЫХ РАСХОДОВ, ВЛЕКУЩИХ ПРИМЕНЕНИЕ КЛАССИЧЕСКИХ КОДОВ С ОБНАРУЖЕНИЕМ И ИСПРАВЛЕНИЕМ ОШИБОК С ВЫСОКОЙ ИЗБЫТОЧНОСТЬЮ ПРИ АНАЛИЗЕ БИОМЕТРИЧЕСКИХ ДАННЫХ

В. Е. Кузнецов[1], А. В. Безяев[2]

[1] Научно-производственное предприятие «Рубин», г. Пенза
[2] Пензенский филиал Научно-технического центра «Атлас», г. Пенза

Аннотация. Предложено упростить анализ энергозатрат и анализ накладных расходов на аппаратно-программную реализацию циклических кодов с высокой избыточностью через их представление линейными рекуррентами (цепь повторителей со сложением по модулю два в обратной связи). Формально этот упрощенный способ декомпозиции основной задачи позволяет легко оценить длину последовательности необходимого числа логических элементов и числа операций сложения по модулю два. Как результат мы имеет оценку минимальных аппаратно-программных затрат и вторую максимальную оценку без учета положения точек подключения сумматоров по модулю два в линейной обратной связи рекуррентны.

Ключевые слова: циклические коды, коды с обнаружением и исправлением ошибок, биометрические данные

ASSESSMENT OF HARDWARE AND ENERGY OVERHEAD COSTS RESULTING IN THE APPLICATION OF CLASSIC CODES WITH ERROR DETECTION AND CORRECTION WITH HIGH REDUNDANCY IN THE ANALYSIS OF BIOMETRIC DATA

V. E. Kuznetsov[1], A. V. Bezyaev[2]

[1] Scientific-industrial Enterprise «Rubin», Penza
[2] Penza branch of Scientific and Technical Center «Atlas», Penza

Abstract. It is proposed to simplify the analysis of energy consumption and the analysis of overhead costs for the hardware and software implementation of cyclic codes with high redundancy, through their representation by linear recurrent circuits (a chain of repeaters with modulo two addition in feedback). Formally, this simplified method of decomposing the main problem makes it possible to easily estimate the length of the sequence of the required number of logical elements and the number of addition operations modulo two. As a result, we have an estimate of the minimum hardware and software costs and a second maximum

estimate without taking into account the position of the adder connection points modulo two in the recurrent linear feedback.
Keywords: cyclic codes, error detection and correction codes, biometric data

В начале 21 века исследователи США и иных стран НАТО активно вели работы по применению классических кодов с обнаружением и исправлением ошибок «сырых» биометрических данных [1, 2, 3, 4]. К сожалению, «сырые» коды реальных биометрических образов имеют от 20 % до 30 % ошибочных бит. Как результат для обнаружения и исправления ошибок в «сырых» данных приходится использовать классические коды с высокой избыточностью. Так при анализе рисунка радужной оболочки глаза используются циклические коды БЧХ с входными данными 2048 бит и длинной выходного скорректированного кода ключа порядка 128 бит [3]. То есть, описанная в работе [3] кодовая конструкция имеет почти 20-ти кратную кодовую избыточность (без учета влияния маскирования наиболее нестабильных разрядов).

В тот момент, когда создавались и исследовались подобные кодовые конструкции была ориентация их реализации с привлечением аппаратно-программных средств достаточной мощности. На текущий момент перспективные биометрические технологии должны быть ориентированы реализацию в доверенном контроллере SIM-карты, RFID-карты. Такие контроллеры имеют малое энергопотребление, низкую производительность, ограниченный объем оперативной и долговременной памяти.

В связи с высокой энергоемкостью и необходимостью сложной программной реализации классических самокорректирующихся кодов возникает задача поиска альтернативного решения задачи, например, с использованием нейронных сетей, автоматически обучаемых по ГОСТ Р 52633.5 [5] и компактных кодов обнаружения и исправления ошибок без накладных расходов на избыточность [6].

Следует отметить, что реализовывать программно классические самокорректирующиеся коды с высокой избыточностью для проведения численных экспериментов является достаточно трудоемкой задачей. Гораздо более простой задачей является программирование рекуррент или CRC средств подсчета контрольных сумм. На рис. 1 представлена структура CRC-4 подсчета контрольных сумм, реализованная на базе последовательности пяти логических элементов (повторителей), ожвеченных петлей обратной связи с двумя сумматорами по модулю два – ⊕.

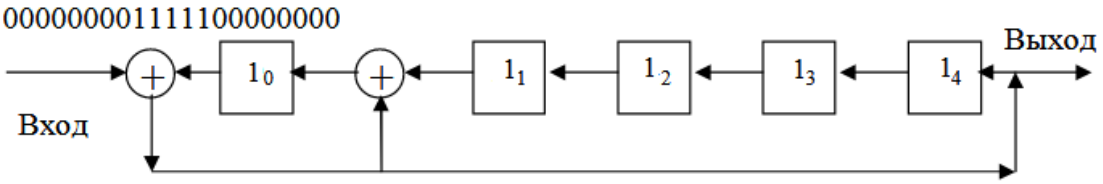


Рис. 1. Линейная рекуррента CRC-4 подсчета контрольных сумм, сворачивающая (хэширующая) любую входную последовательность до 5 бит контрольной суммы

Очевидно, что при необходимости могут быть использованы и более сложные рекурренты, построенные на стандартах подсчета контрольных сумм CRC-8, CRC-16, CRC-32, CRC-64. При этом для их аппаратно-реализации реализации потребуется 8,16, 32, 64 логических элемента и существенно меньшее число отводов обратных связей с логическим элементом, выполняющим операцию сложения по модулю два.

На рис. 2 приведена таблица примитивных полиномов, опираясь на которые формируются отводы обратной связи рекуррент по международным стандарта CRC-N с учетом числа обратных связей и, соответственно, числа необходимых операций сложения по модулю два.

Биты, n	Примитивный многочлен	Период, $2^n - 1$	Число примитивных многочленов	Число $\oplus$
2	$x^2 + x + 1$	3	1	2
3	$x^3 + x^2 + 1$	7	2	2
4	$x^4 + x^3 + 1$	15	2	2
5	$x^5 + x^3 + 1$	31	6	2
6	$x^6 + x^5 + 1$	63	6	2
7	$x^7 + x^6 + 1$	127	18	2
8	$x^8 + x^6 + x^5 + x^4 + 1$	255	16	4
9	$x^9 + x^5 + 1$	511	48	2
10	$x^{10} + x^7 + 1$	1023	60	2
11	$x^{11} + x^9 + 1$	2047	176	2
12	$x^{12} + x^{11} + x^{10} + x^4 + 1$	4095	144	4
13	$x^{13} + x^{12} + x^{11} + x^8 + 1$	8191	630	4
14	$x^{14} + x^{13} + x^{12} + x^2 + 1$	16383	756	4
15	$x^{15} + x^{14} + 1$	32767	1800	2
16	$x^{16} + x^{14} + x^{13} + x^{11} + 1$	65535	2048	4

Рис. 2. Примитивные многочлены для реализации рекуррент подсчета контрольных сумм от CRC-2 до CRC-16 с указанием необходимого числа операций сложения по модулю два

Как видно из таблицы полином CRC-4 (строка 3) должен иметь 5 элементов задержки, два отвода обратных связей и два логических элемента сложения по модулю два. Таким образом, зная порядок рекурренты мы легко можем предсказать число логических элементов, число связей, число операций сложения по модулю два.

В частности, если Даугман [3], решится на то, чтобы только обнаруживать ошибки в выходном коде длинной 2048 бит, то ему потребуется минимальные затраты на реализацию рекурреты CRC-128. В этом случае его «нечеткий экстрактор» будет с очень высокой вероятностью 2^{1920} будет обнаруживать наличие ошибок в выходном криптографическим ключе.

Если Даугман [3], будет использовать самокорректирующий код БЧХ длинной входных данных 2048 бит, а выходных данных 128 бит, то ему потребуется увеличить число элементов для реализации CRC-2048 рекурренты. Придется увеличить число логических элементов как минимум в 16 раз. Если предположить, что число ответвлений обратных связей и число элементов сложения по модулю два растут пропорционально, то мы при программировании самокорректирующегося кода в контроллере (в программируемой логической матрице) будем иметь 16-ти кратное аппаратное усложнение, 16-ти кратное снижение времени вычислений, 16-ти кратный рост потребление энергии. Все это свидетельствует о низкой эффективности классических кодов способных обнаруживать и исправлять ошибки.

Положение усугубляется тем, что приложения нейросетевого искусственного интеллекта в ближайшем будущем предполагается защищать средствами гомоморфного шифрования [6, 8]. Любое средство гомоморфного шифрования само способно накапливать ошибки и является еще одним самостоятельным источником ошибок в кодах. Актуальность задачи синтеза эффективных средств нейросетевого обнаружения и исправления ошибок увеличивается [9].

Список литературы

1. Monrose F., Reiter M., Li Q., Wetzel S. Cryptographic key generation from voice // In Proc. IEEE Symp. on Security and Privacy. Oakland, CA, USA, 2001. P. 202–213.

2. Dodis Y., Reyzin L., Smith A. Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy // In EUROCRYPT. 2004. P. 523–540.

3. Feng Hao, Ross Anderson, and John Daugman. Crypto with Biometrics Effectively // IEEE Transactions on computers. 2006. Vol. 55, № 9.

4. Чмора А. Л. Маскировка ключа с помощью биометрии // Проблемы передачи информации. 2011. № 2 (47). С. 128–143.

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

6. 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. ISBN 978-5-907262-59-1

7. Безяев А. В., Иванов А. И., Корнеев О. В. Типовая схема защиты нейросетевых архивов биометрических данных не криптографическим хешированием через применение линейной рекурренты подсчета контрольных сумм CRC-4 // Труды научно-технической конференции кластера пензенских предприятий, обеспечивающих безопасность информационных технологий. Пенза, 2016. Т. 10. С. 15–20. URL: http://пниэи.рф/activity/science/BIT/T10-p15.pdf

8. Князьков В. С., Исупов К. С., Иванов А. И., Артемов И. И. Гибридные вычисления и их применение для построения гомоморфных нейросетевых систем доверенного искусственного интеллекта // Наука, общество, технологии: проблемы и перспективы взаимодействия в современном мире : монография / Э. М. Абакирова [и др.]. Петрозаводск : МЦНП «Новая наука», 2022. 438 с. ISBN 978-5-00174-630-0

9. Иванов А. И., Иванов А. П., Макарычев П. П. [и др.]. Рост корректирующей способности нейросетевых конструкций с избыточностью за счет замены в них бинарных нейронов на троичные нейроны // Известие вестник учебных заведений. Поволжский регион. Технические науки. 2022. № 3. С. 27–36.

УДК 004.056

КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ УГРОЗ НЕСАНКЦИОНИРОВАННОГО ДОСТУПА К ПЕРСОНАЛЬНЫМ ДАННЫМ АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ ОТДЕЛА КАДРОВ

С. Е. Кондаков[1], К. С. Чудин[2]

[1,2] Московский государственный технический университет имени Н. Э. Баумана, г. Москва

Аннотация. Рассматриваются основы построения концептуальной модели угроз несанкционированного доступа к персональным данным при автоматизации деятельности отдела кадров по отношению к лицам высокого доверия. Описывается порядок построения концептуальной модели, его основные этапы, обоснованы допущения и ограничения. Проведена формальная интерпретация действий нарушителя. Рассмотрены и описаны основные этапы формирования концептуальной модели для получения исследовательского аппарата для оценки эффективности мер обеспечения защиты персональных данных.

Ключевые слова: защита персональных данных, концептуальная модель, угрозы безопасности

CONCEPTUAL MODEL OF THREATS OF UNAUTHORIZED ACCESS TO PERSONAL DATA OF AN AUTOMATED SYSTEM OF THE HR DEPARTMENT

S. E. Kondakov[1], K. S. Chudin[2]

[1,2] Moscow State Technical University named after N. E. Bauman, Moscow

Abstract. The article discusses the basics of constructing a conceptual model of threats of unauthorized access to personal data (hereinafter referred to as PD) when automating the activities of the HR department in relation to high-trust individuals. The procedure for constructing a conceptual model, its main stages are described, assumptions and limitations are justified. A formal interpretation of the offender's actions was carried out. The main stages of the formation of a conceptual model for obtaining a research apparatus for assessing the effectiveness of measures to ensure the protection of personal data are considered and described.

Keywords: personal data protection, conceptual model, security threats

Актуальность

Защита персональных данных (ПДн) – это многогранная область знаний и деятельности, включающая основные свойства систем ПДн с точки зрения их безопасного использования, причины и виды нарушения этих свойств, субъектов информационных отношений, нарушающих безопасное использование систем, методы и средства противодействия нарушителям. Практика показывает, что для анализа такого значительного набора факторов, объектов и действий целесообразно использовать методы моделирования, при которых формируется формальное представление проблемной области на понятийном уровне, т.е. концептуальной модели угроз [1].

Таким образом, актуальной задачей для наглядного представления проблем информационной безопасности является построение концептуальной модели угроз НСД к ПДн при создании автоматизированной системы отдела кадров (далее АС ОК), которая включает основные понятия и определения в предметной области, а также взаимосвязь этих понятий.

Концептуальная модель угроз информации в АС ОК должна соответствовать принятой политике безопасности и является её вербальным описанием, которое содержит и описывает наиболее значимые компоненты (рисунок 1):

– защищаемые ресурсы АС ОК;

– полный перечень дестабилизирующих факторов (ДФ) (модель угроз и модель поведения потенциального нарушителя);

– средства и меры защиты информации, используемые в АС ОК.

Рис. 1. Концептуальная модель защиты информации в АС отдела кадров

Основным объектом защиты является информация, циркулирующая и хранимая в АС ОК (в виде данных, команд, сообщений и т.д.), при этом обеспечивается ее конфиденциальность, доступность и целостность, т.е. предотвращается утечка, хищение, утрата, искажение, подделка, копирование, блокирование информации, а также другие формы несанкционированного вмешательства в информационные ресурсы. Информация в АС ОК имеет определенную цену для предприятия, содержащего кадровую службу, и потенциального нарушителя [2].

В связи с необходимостью организации защиты информации АС ОК по учету информации о лицах, обладающих высоким уровнем доверия, должна развертываться на сертифицированных вычислительных машинах, каждая из которых имеет аппаратно-программное средство ограничения доступа отечественных производителей. Могут быть использованы такие отечественные продукты как «Соболь», «Аккорд-АМДЗ», «Криптон-Замок» и «Витязь» [3].

Факторы [2], воздействующие на защищаемую информацию в процессе функционирования АС ОК, это явления, действия или процессы, результатом которых могут быть искажение, уничтожение защищаемой информации, блокирование доступа к ней. Дестабилизирующие факторы – потенциально существующая совокупность условий, направленных на нарушение штатного режима применения системы с нанесением ей неприемлемого ущерба.

Наиболее серьезными нарушения информационной отдела кадров при реализации им функций передачи, обработки и накопления информации является ее несанкционированное копирование и модификация, а также нарушения доступа к ПДн. Такого рода угрозы безопасности информации не позволяют отделу кадров выполнять свою целевую функцию. Исходя из содержания данной целевой функции, ее невыполнение может иметь серьезные последствия.

Это обусловливает высокие требования к мерам противодействия такого рода угрозам, исключающим возможность нарушения конфиденциальности, целостности и доступности информации в АС отдела кадров крупных предприятий, ориентированных на выполнение государственных заказов.

Концептуальная модель нарушения безопасности информации отдела кадров определяет субъект нарушения, его квалификацию, мотивацию, цели, этапы действий и временные рамки. Подобная модель рассматривается как предпосылка к формальной

интерпретации действий нарушителя. В общем виде содержание концептуальной модели нарушения сводится к следующему:

1. Субъектом нарушения безопасности информации АС ОК является внутренний нарушитель.

2. Действия нарушителя, в том числе и действия в отношении информационных ресурсов и информационных процессов АС ОК, являются противоправными, а сами нарушители квалифицируются как злоумышленники.

3. Имеет место целевая мотивация такого рода действий.

4. Целевой функцией нарушителя является несанкционированное копирование и модификация информации, а также блокирование доступа к ней в АС ОК.

5. Целевая функция нарушителя реализуются в шесть этапов:

– этап преодоления механизмов защиты информации от НСД в АС ОК (этап 1);

– этап регистрации нарушителя как легитимного субъекта доступа к АС (этап 2, например, через сговор с системным администратором средств ограничения доступа);

– этап нарушения конфиденциальности ПДн (этап 3);

– этап нарушения целостности ПДн (этап 4);

– этап нарушения доступности ПДн (этап 5);

– этап скрытия следов несанкционированного воздействия на ПДн (этап 6, например, через сговор с системным администратором средств аппаратно-программной защиты).

6. Соотношения между моментом времени начала действий нарушителя и моментом времени их обнаружения следующие:

– для этапов 1 и 5 характерно практически мгновенное обнаружение действий нарушителя;

– для остальных этапов обнаружение действий нарушителя происходит спустя определенный, иногда значительный, промежуток времени.

7. Для нарушителя характерно многократное (за исследуемый период) выполнение противоправных действий. При этом кратность несанкционированного доступа нарушителем к информации АС ОК определяется его возможностями по обеспечению скрытности своих действий. Вероятность многократного несанкционированного доступа к информации с увеличением кратности существенно снижается, даже в случае сговора с системным администратором средств аппаратно-программной защиты.

Угрозы нарушения конфиденциальности, целостности и доступности информации в АС ОК реализуются нарушителем путем:

– использования штатных компонент операционной среды (ОС) или прикладных программ общего назначения;

– создания нештатных режимов работы программных (программно-аппаратных) средств за счет преднамеренных изменений служебных данных, приводящих к нарушению конфиденциальности, целостности и доступности информации в АС ОК (сговор с системным администратором);

– внедрение вредоносного программного обеспечения (ВПО) с целью блокирования информационных процессов в АС ОК исключено, если по политике информационной безопасности используются только учтенные носители информации.

Рассмотрим реализацию первого этапа рассмотренной ранее последовательности решения задачи исследования, связанного с разработкой функциональных моделей угроз НСД к ПДн в АС ОК и процессов реагирования на такого рода угрозы.

В соответствии с концептуальными положениями методологии системного анализа первоначальными процедурами решения любых задач, связанных с оценкой характеристик исследуемых процессов, включая задачи оценки безопасности ПДн, является формализованное представление этих процессов.

С учетом функционального характера угроз НСД к ПДн в деятельности АС ОК и процессов реагирования на такого рода угрозы в качестве аппарата формализованного представления этих процессов возможно использования методов функционального моделирования. Основу такой возможности составляют три положения, характерных только функциональному представлению этих процессов [4]. Это:

– существование признаков распознавания угроз НСД к ПДн в деятельности АС ОК;

– существование причинно-следственных связей между реализуемыми функциями НСД к ПДн в деятельности АС ОК и реагирования на такого рода угрозы;

– детерминированность порядка выполнения этих функций.

В результате анализа закономерностей практики защиты от такого рода угроз сформировано множество признаков их распознавания, а также установлены варианты порядка реализуемых при этом действий [5].

Множество признаков угроз НСД к ПДн в деятельности АС ОК является той исследовательской средой, на основе которой строится функциональная модель действий нарушителя по реализации такого рода угроз [6]. На основе данного множества формируется множество $Ф^{(1)}$ функций начального (первого) уровня иерархической структуры данной модели. Исходя из композиционного характера построения функциональной структуры описания действий нарушителя по реализации угроз НСД к ПДн в деятельности АС ОК при формировании данного уровня будем исходить из того, что каждой функции множества $H^{(1)}$ будет соответствовать один конкретный признак ее выполнения, т.е. имеет место равенство числа признаков мощности $|H^{(1)}|$ множества $H^{(1)}$ этих функций [7].

На основе множества функций $H^{(1)} = \{н_i^{(1)}\}$, $i = 1, 2, \ldots, I$, первого уровня композиционной структуры функциональной модели действий нарушителя по реализации угроз НСД к ПДн в деятельности АС ОК формируется множество $H^{(2)} = \{н_j^{(2)}\}, j = 1, 2, \ldots, |H^{(2)}|$, функций второго уровня данной структуры данной модели путем установления соответствий между подмножествами множества $H^{(1)}$ и элементами множества $H^{(2)}$.

При этом ряд функций $н_j^{(2)}$ множества $H^{(2)}$, могут не являться композиционно образуемыми, т.е. такими которым будет соответствовать лишь одна, а не подмножество функций множества $H^{(1)}$.

Для множеств $H^{(1)}$ и $H^{(2)}$ будет справедливым неравенство: $|H^{(2)}| < |H^{(1)}|$.

На основе множества функций $H^{(2)} = \{н_j^{(2)}\}, j = 1, 2, \ldots, |H^{(2)}|$, второго уровня композиционной структуры функциональной модели действий нарушителя по реализации угроз НСД к ПДн в деятельности АС ОК формируется множество $H^{(3)} = \{н_k^{(3)}\}$, $k = 1, 2, \ldots, |H^{(3)}|$, функций третьего уровня данной структуры путем установления соответствий между подмножествами множества $H^{(2)}$ и элементами множества $H^{(3)}$.

Как и в случае формирования множества $H^{(2)}$ на основе множества $H^{(1)}$ ряд функций $н_k^{(3)}$ множества $H^{(3)}$, могут не являться композиционно образуемыми.

Для множеств $H^{(2)}$ и $H^{(3)}$ будет справедливым неравенство: $|H^{(3)}| < |H^{(2)}|$.

Формирование композиционной структуры функциональной модели действий нарушителя по реализации угроз НСД к ПДн

в деятельности АС ОК осуществляется до образования целевой функции $H^{(N)}$ – «НСД к ПДн в деятельности АС ОК» – функции N-го (верхнего) уровня композиционной структуры функциональной модели.

Полученное в результате функциональной композиции структурное представление целевой функции $H^{(N)}$ – «НСД к ПДн в деятельности АС ОК» обладает существенным недостатком – индексация функций в структуре не отражает их иерархическую взаимосвязь. Для устранения данного недостатка осуществляется преобразование индексов функций из их композиционного (порядкового) представления в декомпозиционное (иерархическое) [7].

В общем виде функцию исходного уровня рассматриваемой функциональной структуры как элемент композиционного базиса функциональной модели действий нарушителя по реализации угрозы НСД к ПДн в деятельности АС ОК представим как: $y_{g,\,h,\,\ldots,\,u,\,x,\,z}$, где g, h, …, u, x, z – номера функций N-1-го, N-2-го, …, третьего, второго и первого уровней, соответственно, которым принадлежит данная функция, z – ее текущий номер в составе x-й функции второго уровня. При этом, число цифр в индексе $q(n)$ и номер уровня n связаны соотношением:

$$n = N - q(n) \qquad (1)$$

В терминах представленной индексации функции второго, третьего, … и N–1-го уровней рассматриваемой композиционной структуры представляются как $y_{g,\,h,\,\ldots,\,x,\,u}$, $y_{g,\,h,\,\ldots,\,x}$, … и y_g соответственно, а функцию N-го уровня (целевую функцию) – как y.

Исходя из соответствий между функциональным представлением действий нарушителя по реализации угроз НСД к ПДн в деятельности АС ОК и функциональным описанием процесса реагирования на такого рода угрозы, декомпозиционную структуру последнего представим в аналогичном виде. С этой целью обозначим через $з$ целевую функцию «Защита ПД от НСД в деятельности АС ОК», а композиционно связанные множества элементов данной целевой функции (с учетом уровней их декомпозиционного представления) через $з_{g,\,h,\,\ldots,\,x}$, $з_{g,\,h,\,\ldots,\,x,\,u}$ и $з_{g,\,h,\,\ldots,\,x,\,u,\,z}$.

При этом имеют место следующие соответствия между элементами структуры целевой функции y – «НСД к ПДн в деятельности АС ОК» и целевой функции $з$ – «Защита ПД от НСД в деятельности АС ОК».

Таким образом, в статье рассмотрены основы построения концептуальной модели угроз к персональным данным АС ОК, обоснованы допущения и ограничения. Проведена формальная интерпретации действий нарушителя с описанием основных этапов формирования концептуальной модели для получения исследовательского аппарата для оценки эффективности мер обеспечения защиты ПДн в деятельности АС ОК.

Список литературы

1. Кондаков С. Е., Салин В. А. Обобщенный алгоритм проведения исследований по количественной оценке уровня защищенности объекта ИРС специального назначения. М., 2002. 210 с.

2. Кондаков С. Е., Скрыль С. В. Постановка задачи обоснования варианта системы защиты информации комплекса средств автоматизации информационно-расчетной системы ракетных войск стратегического назначения // Научно-технический сборник ОАО «Концерн «Системпром». 2014. № 1 (5). С. 17–22.

3. Саяркин Л. А., Зайцева А. А., Лапин С. П., Домбровский Я. А. Программно-аппаратные средства защиты автоматизированных систем от несанкционированного доступа // Молодой ученый. 2017. № 13 (147). С. 19–22.

4. Скрыль С. В., Шелупанов А. А. Основы системного анализа в защите информации : учеб. пособие для студентов высших учебных заведений. М. : Машиностроение, 2008. 138 с.

5. Кондаков С. Е., Рассохин Г. Н. Функционал качества для выбора варианта АПК АИС с показателями различной природы, размерности и вектора полезности // Научно-технический сборник № 3, Юбилейный: 4 ЦНИИ МО РФ. М., 2014. С. 6–11.

6. Скрыль С. В., Кондаков С. Е., Чудин К. С. Обоснование показателя для оценки эффективности мер обеспечения защиты персональных данных в деятельности кадрового органа службы защиты государственной тайны // Информационная безопасность – актуальная проблема современности. Совершенствование образовательных технологий подготовки специалистов в области информационной безопасности : материалы XXI Всерос. межведомственной науч.-техн. конф. Т. 1. Краснодар : КВВУ, 2020. С. 19–24.

7. Скрыль С. В., Малышев А. А., Волкова С. Н., Герасимов А. А. Функциональное моделирование как методология исследования конфиденциальности информационной деятельности // Интеллектуальные системы : тр. IX Междунар. симп. М. : РУСАКИ, 2010. С. 590–593.

УДК 004.056

РАЗРАБОТКА ЗАЩИЩЕННОЙ КОРПОРАТИВНОЙ СИСТЕМЫ ОБМЕНА СООБЩЕНИЯМИ

А. Н. Крутов

Самарский национальный исследовательский университет имени академика С. П. Королева, г. Самара

Аннотация. Предложен вариант реализации защищенной корпоративной системы для обмена сообщениями. Описана структура приложения и схема безопасности.

Ключевые слова: защищенная корпоративная система, шифрование, анализ защищенности

DEVELOPMENT OF A SECURE CORPORATE MESSAGING SYSTEM

A. N. Krutov

Samara National Research University named after Academician S. P. Korolev, Samara

Abstract. A variant of the implementation of a secure corporate messaging system is proposed. The structure of the application and the security scheme are described.

Keywords: secure corporate system, encryption, security analysis

В современном мире практически все компании вместе с электронной почтой и сотовой связью используют альтернативные каналы связи для решения рабочих вопросов. Такими каналами могут быть приложения для видеозвонков, мессенджеры, чаты, а также различные социальные сети. Однако такие способы коммуникация имеют ряд недостатков, например:

– присутствие контактов сторонних лиц, которые не связаны с рабочим процессом и могут препятствовать выполнению определённых задач;

– под угрозу ставится безопасность компании из-за возможной утечки конфиденциальной информации.

В целях безопасности и удобства коммуникации используются корпоративные системы обмена сообщениями. В рамках компании такие мессенджеры имеют ряд преимуществ над другими альтернативными каналами связи:

– удобство и повышенная скорость взаимодействия между сотрудниками;

– практичность и сосредоточенность на определённых задачах, поставленных компанией.

В настоящее время для крупных организаций представляется целесообразным создавать собственный корпоративный мессенджер, чем пользоваться аналогами. Такое решение обусловлено тем, что иметь собственное приложение более выгодно и более вариативно, так как его функционал может быть сконцентрирован на конкретных задачах, которые необходимы для компании, и все аспекты безопасности могут быть соблюдены.

К главным функциональным требованиям к подобным приложениям относятся быстродействие и корректность работы в аварийных ситуациях, наличие возможности резервного копирования/восстановления, возможность разграничения прав пользователей, путём разделения на уровни доступа к определённой информации. Для обеспечения большей надёжности и защиты от несанкционированного доступа третьих лиц к конфиденциальной информации, которая хранится в БД, применяются средства шифрования данных. Ниже на рис. 1 представлена структура работы приложения.

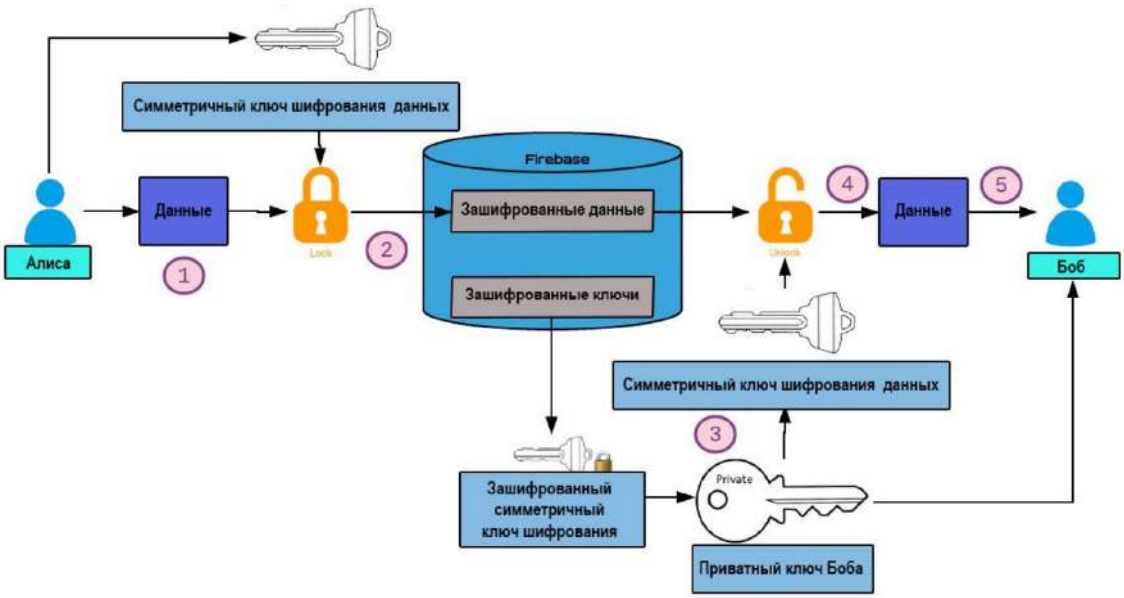


Рис. 1. Структура работы приложения

Приложение работает со встраиваемой реляционной базой данных Firebase. Данная СУБД не использует парадигму клиент-сервер, т.е. ядро СУБД не является отдельно работающим процессом,

с которым взаимодействует программа, а предоставляет библиотеку, с которой программа компонуется. В качестве протокола обмена применяются вызовы функций библиотеки Firebase, что увеличивает быстродействие приложения и упрощает взаимодействие с ним.

Разработанное приложение делится на 4 взаимосвязанных части (модуля):

– модуль аутентификации пользователя;

– модуль регистрации пользователя;

– модуль, содержащий список диалогов;

– модуль диалога.

Так как основными элементами приложения для ОС Android являются активности, то каждая из них соответствует определённому модулю, имеющему свои определённые функции и характеристики в разработанном приложении.

Одним из аспектов защиты от несанкционированного доступа является этап входа пользователя в системе. При входе сотрудником вводятся свои определённые персональные данные, которые необходимы для идентификации его личности. Затем на стороне клиента эти данные шифруются с помощью алгоритма шифрования AES-128 посредством встроенных библиотек Java и передаются по протоколу HTTPS на сервер Firebase. Далее данные проходят проверку на соответствие с информацией, которая хранится в БД и при условии подлинности введённой информации, пользователю предоставляется возможность входа в приложение (рис. 2).

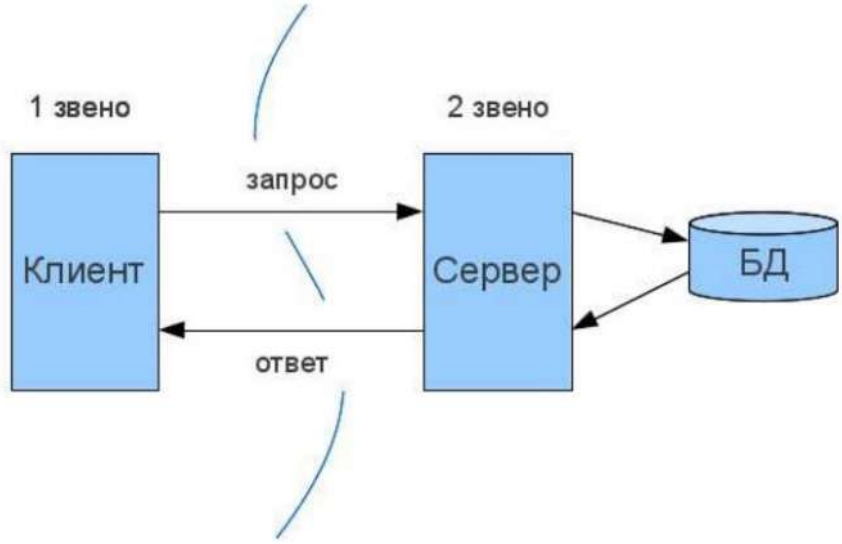


Рис. 2. Схема клиент-серверной архитектуры

Данный способ передачи данных обуславливается тем, что даже при условии компрометации учётной записи администратора БД, никто не сможет посмотреть учётные данные пользователей. Это позволяет обеспечить надёжность хранения данных и повышает уровень информационной безопасности всей компании в целом.

В работе используется двухфакторная авторизация [1]. Так, при входе в аккаунт, пользователю потребуется сначала ввести свой пароль, а дальше код, отправленный по СМС на доверенный номер телефона, указанный при регистрации. В разработанном приложении двухфакторная аутентификация реализуется с помощью специальных встроенных средств FirebasePhoneAuth. Данная опция может находиться как включенном состоянии, так и в выключенном в зависимости от потребности компании в её применении.

Для соблюдения конфиденциальности передаваемой информации в приложении используется шифрование [2]. Данные разработанного приложения, хранящиеся в базе данных, представлены как объекты JSON. Новые объекты превращаются в узлы в структуре, имеющие связанные ключи. Они зашифровываются в пути и хранятся на зашифрованных облачных дисках.

Данные разбиваются на фрагменты файлов для хранения. Каждый фрагмент шифруется на уровне хранилища индивидуальным ключом шифрования. Два фрагмента не имеют одинаковый ключ шифрования, даже если они являются частью одного и того же объекта хранилища. Таким образом, при потенциальной компрометации ключа шифрования данных, компрометация ограничивается только этим фрагментом данных. Это предотвращает доступ к данным без авторизации, повышая безопасность и конфиденциальность данных. Для шифрования данных в разработанном приложении используется алгоритм Advanced Encryption Standard (AES). Все данные на уровне хранилища по умолчанию зашифрованы с помощью AES-128. Шифрование данных в каждом блоке реализуется при помощи ключей шифрования DEK. Для большей безопасности данные ключи шифруются ключевым ключом, имеющим название KEK. Эти KEK хранятся централизованно в KMS-репозитории, созданном специально для хранения ключей (рис. 3). Использование меньшего количества ключей KEK, по отношению к ключам DEK, вместе с применением централизованной службы управления такими ключами, позволяет хранить и шифровать данные с большой управляемостью и безопасностью.

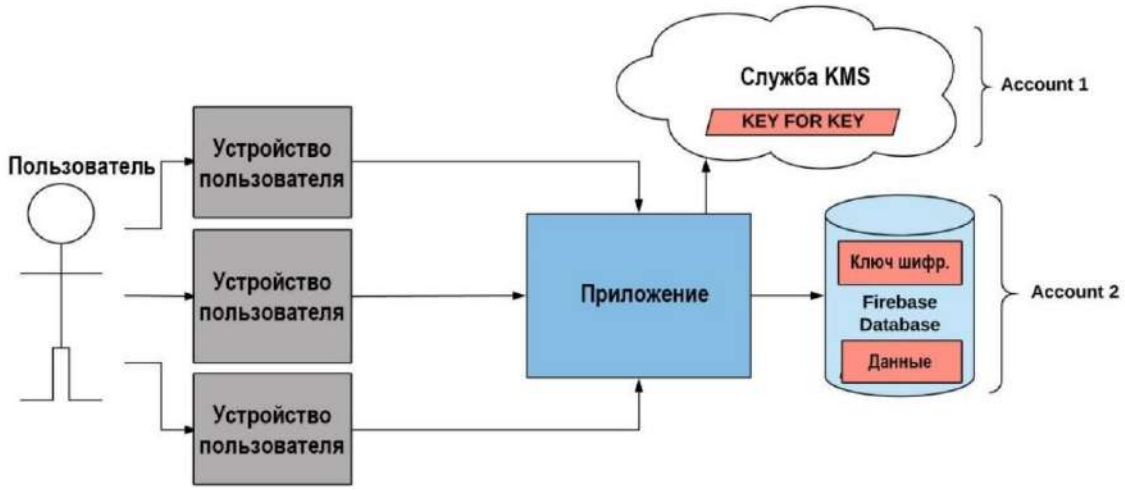


Рис. 3. Схема хранения ключей шифрования DEK и KEK

Благодаря шифрованию у злоумышленника нет возможности получить доступ к данным, случайно завладев ими, не имея доступа к самим ключам. Централизованная служба управления ключами создаёт единую точку, которая обеспечивает и контролирует доступ к ним. Также она занимается своевременной заменой скомпрометированных ключей, что позволяет оперативно предотвратить утечку важной информации. Таким образом, использование представленных выше мер безопасности разработанного приложения позволяет сохранить конфиденциальность пользовательских данных и предотвратить несанкционированный доступ злоумышленников к ним.

Для проверки уровня защищенности использовались методы статического и динамического анализа исходных кодов. При анализе типов безопасности вводится набор типов безопасности (t1, t2, …, tn). Существуют два способа определения типов безопасности всех переменных – ручной и автоматический [3]. Ручной способ обуславливается тем, что при задействовании переменной разработчику необходимо точно определить тип её безопасности. Автоматический способ предполагает, что даны разметка переменных и разметка функций, осуществляющих пользовательский ввод. Во время последовательного анализа программы все переменные, для

Для проведения базового анализа защищённости разработанного приложения определяется список возможных уязвимостей, а также набор инструментов, которыми эти уязвимости будут выявляться. OWASP (Web Application Security Project) – одна из основных методологий тестирования приложений на уязвимости.

Как правило, выделяют основной список уязвимостей, по которым производится сканирование:

– небезопасное хранение данных (Insecure Data Storage);

– небезопасная передача данных (Insecure Communication);

– обход архитектурных ограничений (Improper Platform Usage);

– слабая криптостойкость (Insufficient Cryptography);

– контроль содержимого клиентских приложений и модификация данных (Client Code Qualityand Code Tampering);

– анализ исходного кода (Reverse Engineering).

Для поиска и анализа уязвимостей безопасности использовалось приложение Quixxi. Этот сканер позволяет получить мобильную аналитику, путём предоставления подробной информации в виде отчёта об уязвимостях мобильного приложения, используя методологию OWASP, а также обеспечивает защиту с помощью расширенной версии данного приложения. Проведенный анализ показал, что разработанная корпоративная система обмена сообщениями согласно результатам сканирования и анализа, не имеет уязвимостей высокой критичности.

Список литературы

1. Надейкина В. С., Лагуткина Т. В. Анализ способов реализации системы многофакторной аутентификации // Информационные технологии. 2022. Т. 7, № 4. С. 59–66. doi: 10.18413/2518-1092-2022-7-4-0-7 EDN: EIZSWX

2. Жашкова Т. В., Паршин А. А. Анализ и сравнительная характеристика методов шифрования // Современные информационные технологии. 2022. № 36 (36). С. 126–129. doi: 10.46548/CIT-2022-0036-0031 EDN: RTIIBY

3. Оношко Д. Е., Бахтизин В. В. Метод оценки качества web-приложений, основанный на обнаружении уязвимостей // Цифровая трансформация. 2018. № 1. С. 58–65. EDN: YWWHBI

УДК 004.056

МАСКИРОВАНИЕ ОБРАБАТЫВАЕМОЙ ИНФОРМАЦИИ СРЕДСТВ КРИПТОГРАФИЧЕСКОЙ ЗАЩИТЫ ИНФОРМАЦИИ

С. П. Хворостухин[1], А. М. Масалов[2]

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

Аннотация. Рассматривается применение маскирования информации, обрабатываемой средствами криптографической защиты, для предотвращения атак по техническим каналам утечки информации. Предлагаются варианты модификации узлов обработки информации для обеспечения маскирования.

Ключевые слова: маскирование, средства криптографической защиты информации, технические каналы утечки информации

MASKING THE PROCESSED INFORMATION OF CRYPTOGRAPHIC DEVICES

S. P. Khvorostukhin[1], A. M. Masalov[2]

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

Abstract. The article discusses the use of masking information processed by cryptographic devices to prevent side-channel attacks. Variants of modification of information processing units to provide camouflage are proposed.

Keywords: masking, means of cryptographic information protection, technical channels of information leakage

В зависимости от природы образования информативного сигнала технические каналы утечки можно разделить на естественные и специально создаваемые. Естественные каналы утечки информации образуются за счет побочных электромагнитных излучений, возникающих при обработке информации, наводок информативных сигналов в линиях электропитания, соединительных линиях и посторонних проводниках. К специально создаваемым каналам утечки относятся каналы, создаваемые путем внедрения электронных устройств перехвата информации (закладных устройств) и путем высокочастотного облучения [1].

Для средств криптографической защиты информации, имеющих интерфейсы подключения к сетям обмена данными, например Ethernet, существует вероятность воздействия на аппаратные узлы обработки со стороны каналов связи. Целью такого воздействия может быть активация недокументированных возможностей программных или аппаратных компонентов либо создание сбоев в работе защитных механизмов. Это может привести как к нарушению корректности работы криптографических алгоритмов, так и к записи криптографических параметров, например ключей, в выходные пакеты данных. Для защиты от данной угрозы предлагается реализовать маскирование получаемой и передаваемой информации по всему тракту прохождения.

Маскирование обрабатываемой техническими средствами (в том числе средствами криптографической защиты) информации реализуется в следующих целях [2, 3]:

– защита от утечки информации по каналу побочных электромагнитных ирлучений и наводок (ПЭМИН);

– защита от активации недокументированных возможностей и закладных устройств аппаратных компонентов технического средства;

– защита от активации недокументированных возможностей микропрограмм активных средств обработки информации (процессоры, микроконтроллеры);

– защита от утечки обрабатываемой информации в результате сбоев.

Реализация защиты от утечки информации по каналу ПЭМИН обеспечивается изменением структуры обрабатываемых информационных блоков и, как следствие, изменением структуры сигналов ПЭМИН.

Реализация защиты от активации недокументированных возможностей и закладных устройств обеспечивается изменением поступающих с канала связи информационных блоков в соответствии с алгоритмом маскирования, что позволяет уменьшить вероятность доведения управляющих последовательностей до целевых компонентов технического атакуемого технического средства.

Реализация защиты от утечки информации в результате сбоев обеспечивается выполнением операций с маскированной информации на всем тракте обработки, таким образом, при ошибочной записи защищаемой информации в выходную последовательность одним из элементов тракта обработки, такая информация

будет замаскирована следующим элементов и попадет в канал связи в измененном виде.

Существуют несколько способов маскирования информации в технических средствах. Основными из них являются [4]:

– динамическое маскирование;

– статическое маскирование;

– маскирование в системах защиты информации;

– криптографическое маскирование;

– матричное маскирование.

Динамическое маскирование основывается на принципе маскирования и снятия маски данных при выполнении обращения к ним. Это означает что запрос пользователя приходит не на целевую систему, а на развернутое на отдельном сервере ПО для маскирования. Динамическое маскирование также может реализоваться в приложениях, которые построены по трехзвенной архитектуре. Она представляет собой установку программного обеспечения для маскирования непосредственно на сервер приложения, что разрешает маскировать и перехватывать данные запросах пользователей в момент их передачи на драйвер без данных.

Статическое маскирование основывается на технической реализации с помощью специализированных программ на отдельном сервере или виртуальной машине. Система выполняет запросы к базе данных и, используя методы СУБД, либо создает копию с измененными данными, либо маскирует непосредственно саму базу данных.

Маскирование в системах защиты информации или системах мониторинга очень схож с динамическим маскированием. Данные маскируются до записи на накопитель информации, т.е. на лету. Это делается чтобы исключить возможность их получения оператором средств защиты информации в незамаскированном виде.

Криптографическое маскирование основывается на маскировании цифровой визуальной информации с применением криптографических методов или же их элементов [5].

Матричное маскирование не используется криптографические подходы, а только лишь матричная арифметика. Данный способ маскирования в настоящее время основывается на широком применении программируемой логики в качестве вычислительной основы встраиваемых распределенных портативных устройств и интегральных схем цифровых сигнальных процессоров [5]. Интегральные схемы данных классы имеют аппаратные модули,

которые ускоряют операцию свертки и, в следствии чего, скалярное произведение векторов. Из этого выходит, что матричное умножение осуществляется на аппаратном уровне, что определяет лучшую производительность реализации вычислителя при одинаковой полупроводниковой технологии производства интегральных схем.

Недостатком матричного маскирования является предложенная матрица (оператор преобразования), которая довольно ресурсоемкая для вычисления, а при этом количество таких матриц очень ограниченно. Однако такого недостатка уже нет у новых уникальных ортогональных базисов, которые включают в себя матрицы Эйлера, Мерсенна, Ферма, Мерсенна-Уолша и другие [6]. Данные матрицы несмотря на то, что включают в себя иррациональные значения, обладают свойством, при котором количество различных значений уровней в общем случае исчисляется единицами.

Предлагается рассмотреть выполнение маскирования и операций с маскированной информацией на примере узла приема-передачи сетевых пакетов и контроллера обработки пакетов, реализуемых в «системе на кристалле» на базе программируемой логической интегральной схемы (ПЛИС). Возможности архитектуры ПЛИС позволяют внедрить элементы подсистемы маскирования информации в исходный код ядер. Реализация подсистемы маскирования приводит к увеличению потребления ресурсов ПЛИС. Временные затраты на выполнение операций должны быть по возможности минимальными. Для реализации в подсистеме маскирования обрабатываемой информации был выбран способ статического маскирования блоком маски, формируемым на сеанс работы СКЗИ. Чтобы маскирование информации не снижало производительности операций по обработке, предлагается осуществлять маскирование в прозрачном режиме – наложение и снятие маски осуществляется в такте выполнения операции, скрытно для узла обработки.

Функциональная схема исходного контроллера приема-передачи сетевых пакетов приведена на рис. 1.

Для реализации функции снятия и наложения маски в контроллер необходимо добавить:

– запоминающее устройство для хранения маски;

– указатель маски в запоминающем устройстве для определения текущего используемого слова маски;

– схему сложения по модулю 2 текущего слова маски и слова данных.

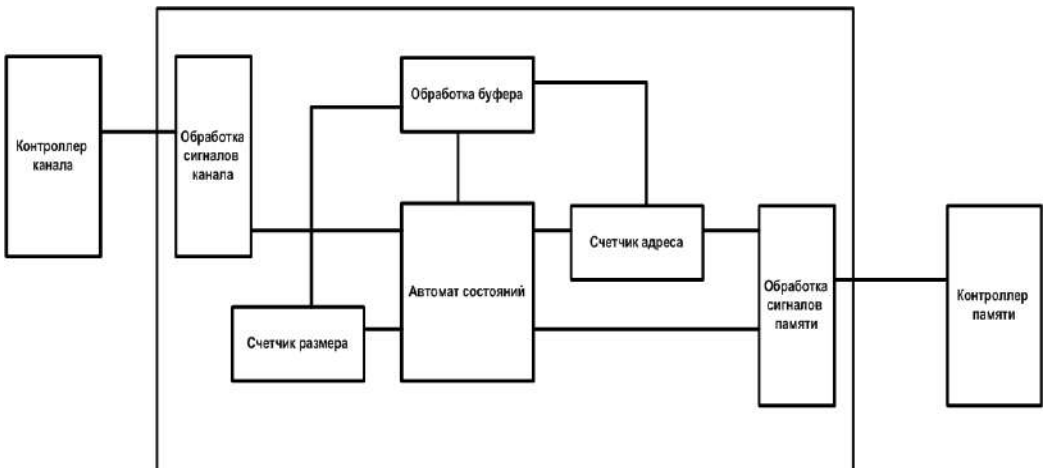


Рис. 1. Функциональная схема исходного контроллера приема-передачи сетевых пакетов

При получении пакета с канала наложение маски выполняется поблочно при записи данных пакета во внутреннюю память. Используемое в каждом такте слово маски определяется счетчиком.

Функциональная схема доработанного контроллера приема-передачи сетевых пакетов приведена на рис. 2.

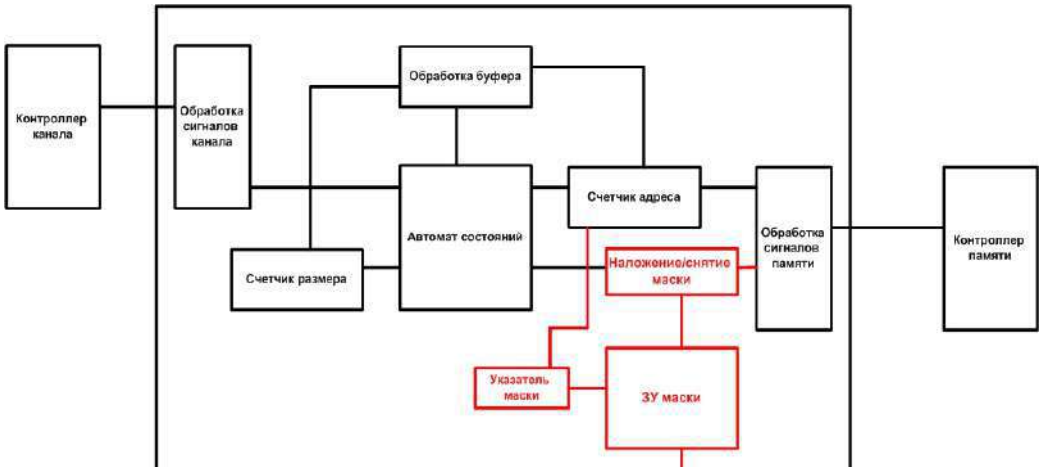


Рис. 2. Функциональная схема исходного контроллера приема-передачи сетевых пакетов

Контроллер обработки пакетов выполняет подготовку пакетов к выполнению криптографических преобразований и обработку после выполнения преобразований. При работе контроллера выполняются операции как с маскированными, так и с немаскированными данными. Для разделения таких операций предлагается выделить массив адресов внутренней памяти контроллера,

для которых операции выполняются с маскированными данным, и массив адресов для работы с немаскируемыми данными. Маскируются только операции ядра контроллера, операции вспомогательных устройств выполняются без маскирования. Операции с маскированными и немаскированными данными должны затрачивать одинаковое количество тактов.

Функциональная схема исходного контроллера обработки приведена на рис. 3.

Рис. 3. Функциональная схема исходного контроллера обработки

Выполнение операций с маскированными данными контроллером обработки выполняется следующим образом. В начале идет ожидание записи защитной маски в контроллер. После записи маски в контроллер, начинается выполнение основной программы. Компоненты наложения/снятия маски осуществляет мониторинг состояний строба выполнения операции ядра контроллера (чтения или записи). Определяется адрес обращения операции. Если адрес попадает в массив маскируемых данных, то выделяется младшая часть адреса обращения. Если нет, то ожидается установка строба выполнения операции с устройствами. Далее выделяется младшая часть адреса обращения и вычисляется из младшей части адреса обращения указатель маски. Извлекается слово маски по вычисленному указателю. Если выполняется операция чтения, то требуется сложить по модулю 2 слово на шине данных памяти и слово маски и записать результат на шину данных памяти. Если операция не выполняется, то складывается по модулю 2 слово на шине данных ядра контроллера и слово маски, и результат записывается на шину данных ядра контроллера.

Функции снятия и наложения маски встраиваются в формирование данных ядра контроллера. Функция выбора слова маски добавляется в систему адресации.

Функциональная схема доработанного контроллера обработки представлена на рис. 4.

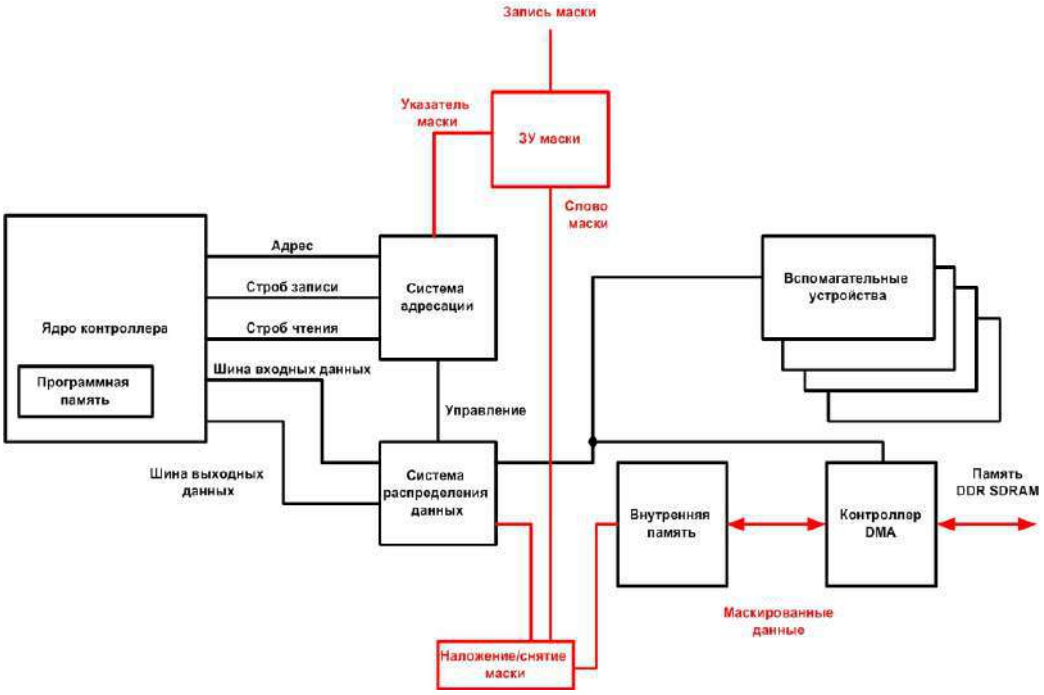


Рис. 4. Функциональная схема доработанного контроллера обработки

Предлагаемые модификации должны быть реализованы в компонентах ПЛИС для обеспечения маскирования обрабатываемой информации, поступающей в сетевых пакетах в качестве защитной меры от утечки по техническим каналам.

Список литературы

1. Хорев А. А. Техническая защита информации : учеб. пособие для студентов вузов : в 3 т. Т. 1: Технически каналы утечки информации. М. : Аналитика, 2008. 436 с.

2. Литвинов М. Ю. Алгоритмы маскирующих преобразований видеоинформации М. ; СПб. гос. ун-т аэрокосм. приборостроения, 2009. 122 с.

3. А. с. № 1773220, кл. 04 К 3/00, Россия. Способ маскировки радиоизлучений средств вычислительной техники и устройство для его реализации / Дмитриев А. С., Залогин Н. Н., Иванов В. П. [и др.]. Приоритет от 21 сентября 1981 г.

4. Маскирование критичных данных: выбираем метод. URL: https://www.anti-malware.ru/practice/solutions/Masking-sensitive-data

5. Востриков А. А., Сергеев М. Б., Литвинов М. Ю. Маскирование цифровой визуальной информации: термин и основные определения // Информационно-управляющие системы. 2015. № 5. 123 с.

6. Балонин Ю. Н., Востриков А. А., Капранова Е. А. [и др.]. Цифровое маскирование матрицами Мерсенна и его особые изображения // Фундаментальные исследования. 2017. № 4-1. С. 13–18.

УДК 004.056

РЕАЛИЗАЦИЯ АЛГОРИТМОВ ПРОВЕРКИ И ФОРМИРОВАНИЯ ЦИФРОВОЙ ПОДПИСИ НА ПЛИС

С. П. Хворостухин[1], Е. К. Самарцев[2]

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

Аннотация. Рассматриваются особенности реализации алгоритмов проверки и формирования цифровой подписи на ПЛИС в виде модуля «системы на кристалле». Приведены результаты моделирования разработанных блоков. Сформированы предложения по дальнейшей оптимизации реализации.

Ключевые слова: цифровая подпись, эллиптические кривые, «система на кристалле»

IMPLEMENTATION OF ALGORITHMS FOR VERIFYING AND FORMING A DIGITAL SIGNATURE ON FPGA

S. P. Khvorostukhin[1], E. K. Samartsev[2]

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

Abstract. The features of the implementation of algorithms for verifying and forming a digital signature on an FPGA in the form of a «system on a chip» module are considered. The results of modeling the developed blocks are presented. Proposals for further optimization of implementation have been formed.

Keywords: digital signature, elliptic curves, «system on a chip»

ПЛИС имеют широкое применение при построении средств криптографической защиты [1, 2, 3]. Как правило, на ПЛИС реализуются функции шифрования/расшифрования, а общесистемные и функции предварительной обработки выполняются процессорами. Электронная подпись – информация в электронной форме, получаемая в результате использования криптографических алгоритмов над подписываемой информацией. Алгоритмы формирования и проверки электронной подписи реализуются с использованием операций в группе точек эллиптической кривой, определенной над конечным простым полем [4]. Использование

эллиптических кривых обеспечивает снижение длины ключа и повышение быстродействия. Основной криптографической операцией в таких алгоритмах является операция скалярного умножения точки эллиптической кривой на целое число, которая реализуется через сложение и удвоение точек эллиптической кривой. Последние в свою очередь выполняются на основе операции сложения, умножения и инвертирования в конечном поле, над которым рассматривается кривая [5]. В средствах криптографической защиты цифровая подпись может быть использована при аутентификации участников информационного обмена, аутентификации источников данных (носителей информации), реализации протоколов открытого распределения ключей, построенных на эллиптических кривых. Возможности ПЛИС в части использования параллельных вычислений и представления чисел большой разрядности позволяют построить высокопроизводительные реализации операций с точками эллиптической кривой, алгоритмов формирования и проверки цифровой подписи.

В отличие от обычных цифровых микросхем, логика работы ПЛИС не определяется при изготовлении, а задаётся посредством программирования. ПЛИС поддерживают многозадачную конфигурацию, где два или более битовых потока конфигурации ПЛИС могут храниться с помощью одного источника конфигурации. Применение ПЛИС контролирует, когда и какую конфигурацию загружать дальше.

«Система на кристалле» (СнК) представляет из себя электронную схему, которая выполняет функции целого устройства и размещена на одной интегральной схеме [6]. Применение СнК, реализуемой на базе ПЛИС, позволяет отказаться от использования микросхем центрального процессора. Все функции, связанные с обработкой и управлением каждым блоком устройства будут сосредоточены в соответствующей микросхеме ПЛИС. Основные функции по управлению СнК возлагаются на «программный» процессор – модуль программного обеспечения ПЛИС, реализующего логику работы центрального процессора [7].

Типичная СнК содержит:

– один или несколько микроконтроллеров, процессоров или ядер цифровой обработки сигналов. СнК содержащую несколько процессоров, называют многопроцессорной системой на кристалле;

– банк памяти, состоящий из модулей ПЗУ, ОЗУ, ППЗУ или флеш;

– источники опорной частоты, например, кварцевые резонаторы и схемы фазовой автоподстройки частоты;

– таймеры, счётчики, цепи задержки после включения;

– блоки, реализующие стандартные интерфейсы для подключения внешних устройств, например, USB, FireWire, Ethernet, USART, SPI;

– блоки цифро-аналоговых и аналого-цифровых преобразователей;

– регуляторы напряжения и стабилизаторы питания.

Проектирование СнК на основе ПЛИС разделяют на два этапа:

– проектирование аппаратного обеспечения;

– проектирование программного обеспечения.

При правильной постановке процесса проектирования, данные этапы могут выполняться параллельно. Проектирование аппаратного обеспечения СнК является наиболее трудоемкой процедурой. Ее можно разделить на следующие этапы:

– определение номенклатуры входящих в состав СнК устройств;

– определение типов внутренних и внешних интерфейсных связей;

– проектирование компонентов системы в соответствии с определенными видами межмодульных связей. Результатом проектирования являются описания на языках описания аппаратных средств, подлежащие последующему синтезу;

– сборка системы в единый проект, назначение диапазонов адресного пространства шин устройствам системы;

– задание ограничений проекта: назначения контактов, временных ограничений, начальных состояний распределенной и блочной памяти на целевой ПЛИС;

– автоматизированные процедуры синтеза, размещение и трассировка аппаратного проекта на ПЛИС;

– верификация аппаратного обеспечения проекта.

Разработка ПО опирается на результаты проектирования аппаратного обеспечения, такие как:

– версии компонентов системы;

– настроечные параметры «по умолчанию»;

– базовые адреса устройств;

– алгоритмы инициализации и самоконтроля;

– алгоритмы обмена информацией с периферийными устройствами.

При построении СнК на ПЛИС целесообразным является выведении трудоемких задач в отдельные специализированные модули, функционирующие параллельно центральному процессору. Модули подключаются к системной шине. Для снижения нагрузки на системную шину и увеличения производительности реализуется буферная памяти внутри модулей. Таким образом, обработка данных передается от центрального процессора к набору специализированных модулей. Задачами центрального процессора при этом остаются распределение данных и постановка задач.

Задачами, реализуемыми в виде отдельных модулей, могут быть функции, связанные с криптографическими преобразованиями данных: шифрование, расшифрование, формирование случайных чисел, вычисление хэш-функций, формирование и проверка цифровых подписей. Криптографические функции могут объединены в один криптографический сопроцессор, либо в несколько, вплоть до реализации каждой функции в виде отдельного модуля. Предлагается реализовать в составе СнК на ПЛИС модуль, реализующий операций проверки и формирования электронной подписи, с возможностью доступа к промежуточным операциям – вычисления хэш-функции, операций с точками эллиптической кривой.

При разработке модуля проверки и формирования цифровых подписей должны быть предусмотрены возможности:

– вычисления значения хэш-функции для записанной в буфер последовательности;

– выполнения операции скалярного умножения для заданных числа и точки эллиптической кривой;

– формирование цифровой подписи для записанной в буфер последовательности и записанного ключа подписи;

– проверку цифровой подписи для записанной в буфер последовательности, записанного значения подписи и записанного ключа проверки подписи;

– выполнение контроля на тестовых примерах для всех операций.

Структура модуля включает:

– блок регистров – набор переменных, организованных на основе распределенной памяти ПЛИС, используемых для настройки работы модуля и задания режима работы;

– буфер записи обрабатываемой последовательности, организованный на основе блочной памяти ПЛИС;

– блок вычисления значения хэш-функции;

– блок скалярного умножения точки эллиптической кривой;

– блок сложения точек эллиптической кривой;

– блок удвоения точек эллиптической кривой;

– блок вычисления мультипликативного обратного по модулю;

– блок формирования цифровой подписи;

– блок проверки цифровой подписи.

Набор регистров, определенный в блоке регистров, позволяет выполнять заданные для модуля операции, в том числе и с введением контрольных значений. Блок регистров включает в себя:

– регистр числа p - модуля эллиптической кривой;

– регистр числа q - порядка циклической подгруппы группы точек эллиптической кривой;

– регистр числа k - случайное (псевдослучайное) целое число, используемое в формировании подписи, используется только при контроле;

– регистр значения хэш-функции - используется при контроле;

– регистр размера последовательности;

– регистр числа d - ключа подписи;

– регистр коэффициента x_p точки эллиптической кривой P;

– регистр коэффициента y_p точки эллиптической кривой P;

– регистр коэффициента a эллиптической кривой;

– регистр коэффициента b эллиптической кривой;

– регистр коэффициента x_c точки эллиптической кривой C;

– регистр коэффициента y_c точки эллиптической кривой C;

– регистр коэффициента x_q точки эллиптической кривой Q - ключа проверки подписи;

– регистр коэффициента y_q точки эллиптической кривой Q - ключа проверки подписи;

– регистр блока данных - адрес записи данных во входной буфер.

Структура блока регистров модуля приведена на рис. 1.

Реализация базовых математических функций построена на известных алгоритмах: нахождения мультипликативного обратного по модулю реализуется через расширенный алгоритм Евклида, умножение и деление чисел через классические алгоритмы [5].

Рис. 1. Структура блока регистров разработанного модуля

По результатам моделирования в среде ISim, были получены первичные результаты по производительности разрабатываемого модуля. Моделирование выполнялось на тестовых примерах, приведенных в ГОСТ [4] для размерности ключа подписи 256 бит. Моделирование выполнялось для тактовой частоты работы модулей 100МГц. Результаты моделирования представлены на рис. 2 и рис. 3. Время, затраченное в модели для формирования подписи по предварительно загруженным тестовым данным, составило 317 мс, для проверки подписи потребовалось 632 мс.

Рис. 2. Результаты моделирования процесса формирования цифровой подписи на тестовом примере

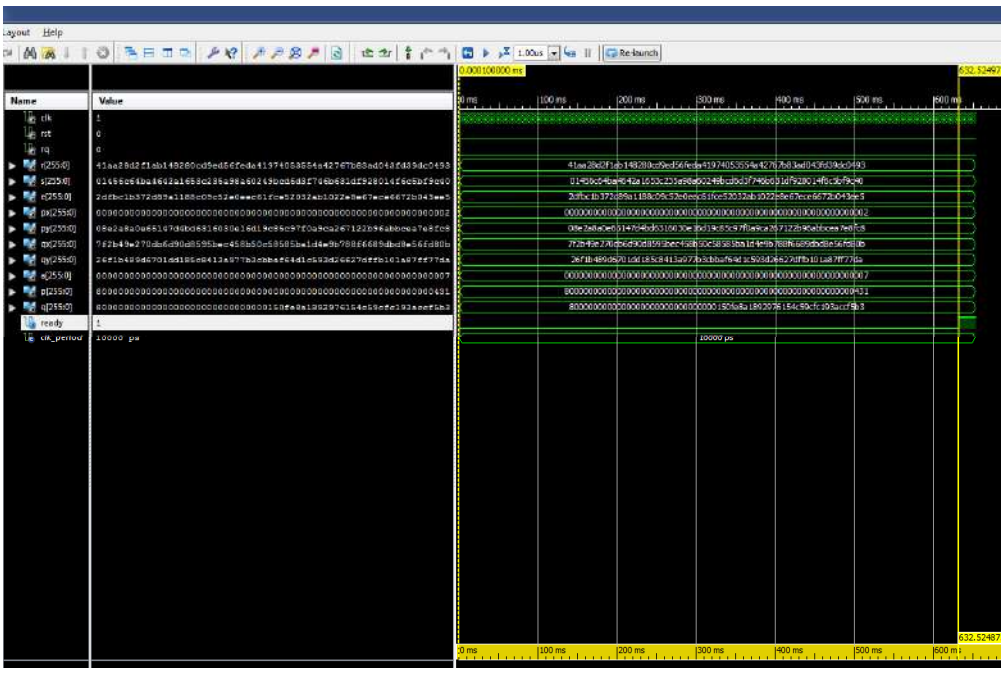


Рис. 3. Результаты моделирования процесса проверки цифровой подписи на тестовом примере

Разрабатываемый модуль требует дальнейшей оптимизации, при этом направлениями оптимизации могут быть как уменьшение числа используемых ресурсов ПЛИС, так и повышение производительности [8].

Для уменьшения количества используемых ресурсов возможным является ограничение количества блоков умножения, модуля, сложения и удвоения точки эллиптической кривой.

Повышение производительности может быть достигнуто распараллеливанием процесса вычисления, однако, это требует введение в состав модуля дополнительных вычислительных блоков.

Отдельной проработки требует возможность перехода от операций с точками эллиптической кривой в форме Вейерштрасса к точкам эллиптической кривой в форме Эдвардса [9].

Список литературы

1. Кожевников А. А. Разработка и исследование высокоскоростных шифраторов на основе ПЛИС Spartan-6 // Техническая кибернетика, радиоэлектроника и системы управления : XI Всерос. науч. конф. молодых учёных, студентов и аспирантов : сб. материалов. Таганрог : Изд-во ЮФУ, 2012. Т. 2. С. 103.

2. Кашаев Е. Д., Хворостухин С. П., Дмитриев М. А., Ильин А. И. Особенности использования ПЛИС в средствах защиты информации //

Многоядерные процессоры, параллельное программирование, ПЛИС, системы обработки сигналов : сб. ст. регион. науч.-практ. конф. (28 февраля 2012 г. на базе Алтайского государственного университета). Барнаул, 2012. С. 98–102.

3. Хворостухин С. П. Последовательно-параллельная организация узла шифрования на базе ПЛИС // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. (Пенза, 03 июня 2020 г.). Пенза : Изд-во ПГУ, 2020. С. 141–145. EDN: AYHRNZ

4. ГОСТ 34.10–2018. Межгосударственный стандарт. Информационная технология. Криптографическая защита информации. Процессы формирования и проверки электронной цифровой подписи : [утв. и введен в действие приказом Федерального агентства по техническому регулированию и метрологии от 4 декабря 2018 г. № 1059-ст].

5. Болотов А. А., Гашков С. Б., Фролов А. Б., Часовских А. А. Элементарное введение в эллиптическую криптографию: алгебраические и алгоритмические основы. М. : КомКнига, 2012. 356 с.

6. Постоев А. И., Соловьев А. А., Иордан В. И., Скобкарев С. А. Использование ПЛИС архитектуры FPGA для проектирования «системы на кристалле» в составе высокоскоростного видеорегистратора потока изображений // Многоядерные процессоры, параллельное программирование, ПЛИС, системы обработки сигналов : сб. ст. регион. науч.-практ. конф. (28 февраля 2012 г. на базе Алтайского государственного университета). Барнаул, 2012. С. 84–89.

7. Барков Е. С., Переверзев А. Л., Силантьев А. М. Высокопроизводительный софт-процессор для встраиваемых систем на основе ПЛИС // Проблемы разработки перспективных микро- и наноэлектронных систем (МЭС). 2022. № 4. С. 130–135. doi: 10.31114/2078-7707-2022-4-130-135. EDN: QLGMKG

8. Акилов А. А., Жарких А. А., Диченко С. А. Анализ построения электронно-цифровой подписи по ГОСТ р 34.10–2012 // Прикладная математика и фундаментальная информатика. 2018. Т. 5, № 3. С. 55–61. doi: 10.25206/2311-4908-2018-5-2-55-61. EDN: YSXFSP

9. Султанов Д. Р. Преимущества применения в микросхемах скрученных эллиптических кривых Эдвардса // Безопасные информационные технологии : сб. тр. Восьмой Всерос. науч.-техн. конф. (Москва, 06–07 декабря 2017 г.) / под ред. М. А. Басараба. М. : Московский государственный технический университет имени Н. Э. Баумана (национальный исследовательский университет), 2017. С. 455–457. EDN: XYSKVV

УДК 621.391:004.032.26

АЛГОРИТМ ВЫДЕЛЕНИЯ ФАЗОВОГО ПУСКА В АППАРАТУРЕ ПЕРЕДАЧИ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ НЕЙРОННОЙ СЕТИ

А. П. Иванов[1], В. Е. Цибин[2], Д. А. Холопов[3]

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

Аннотация. Рассматривается алгоритм выделения сигналов фазового пуска с использованием искусственной нейронной сети. Использование алгоритма в аппаратуре передачи данных позволит обеспечить точное распознавание комбинаций фазового пуска и высокую вероятность их выделения даже при воздействии дестабилизирующих факторов.

Ключевые слова: искусственная нейронная сеть, дестабилизирующие факторы, комбинация фазового пуска

ALGORITHM FOR SEPARATING PHASE START IN DATA TRANSMISSION EQUIPMENT USING A NEURAL NETWORK

A. P. Ivanov[1], V. E. Tsibin[2], D. A. Kholopov[3]

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

Abstract. The article discusses an algorithm for isolating phase start (PT) signals using an artificial neural network. The use of an algorithm in data transmission equipment will ensure accurate recognition of FP combinations and a high probability of their identification even under the influence of destabilizing factors.

Keywords: artificial neural network, destabilizing factors, phase trigger combination

Аппаратура передачи данных (АПД) применяется в автоматизированных системах управления (АСУ) и предназначена для автоматизированного обмена информацией по каналам передачи данных, обеспечения радио- и проводной телефонной связи с вышестоящим пунктом управления, с подчиненными подразделениями и взаимодействующими изделиями [1]. Эффективность функционирования АСУ непосредственно зависит от помехозащищенности АПД, т.е. способности сохранять помехоустойчивость при воздействии различных дестабилизирующих факторов [2].

В настоящее время при передаче данных с использованием радиоканалов в основном используется передача датаграмм, т.е. блоков информации, имеющих определенную структуру, передаваемых протоколом через сеть связи без предварительного

установления соединения и создания виртуального канала. Для передачи датаграмм в АПД используется метод мгновенной установки фазы.

Метод мгновенной установки фазы применяется, когда сообщения передаются эпизодически и каждое из них содержит в своем составе комбинацию ФП, по которой определяется начало информационной части сообщения.

В системах с необходимостью быстрого установления циклового синхронизма в настоящее время для выделения начала информационной части сообщения используются синхронизующие кодовые последовательности. Обычно такие последовательности передаются как часть заголовка сообщения. Приёмник знает эту кодовую последовательность и постоянно ищет её в потоке данных, используя для этого корреляционный метод. В статье предлагается алгоритм для выделения ФП с использованием искусственной нейронной сети, которая предварительно обучена.

Принцип предлагаемого алгоритма выделения сигналов ФП с использованием искусственной нейронной сети показан на рис. 1.

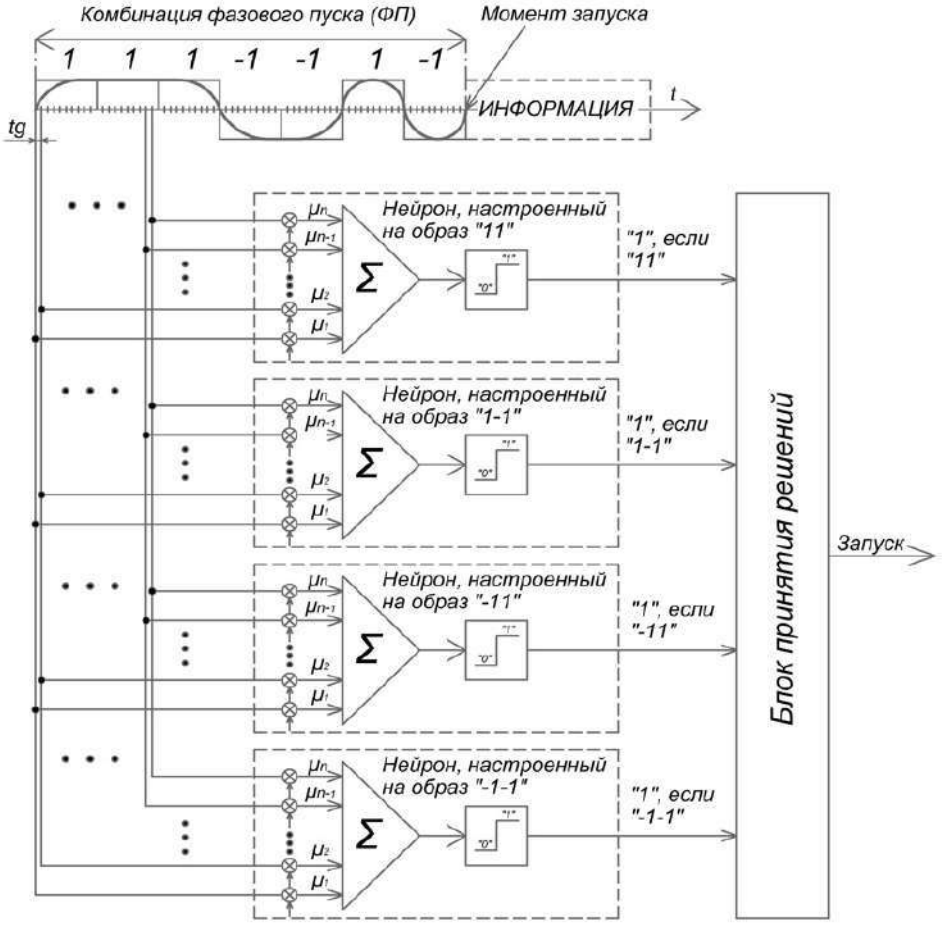


Рис. 1. Алгоритм выделения сигналов ФП с использованием искусственной нейронной сети

Алгоритм заключается в следующем. Сигнал из канала связи через аналогово-цифровой преобразователь поступает на устройство выделения сигналов ФП. В состав устройства выделения сигналов ФП входит искусственная нейронная сеть. Искусственная нейронная сеть представляет собой систему соединенных и взаимодействующих между собой простых процессоров (искусственных нейронов). Каждый процессор подобной сети имеет дело только с сигналами, которые он периодически получает и посылает другим процессорам. Преимущества искусственных нейронных сетей заключаются, во-первых, в распараллеливании обработки информации; во-вторых, способности самообучаться, т.е. создавать обобщения. Под термином «обобщения» понимается способность получать обоснованный результат на основании данных, которые не участвовали в процессе обучения. Технически обучение заключается в нахождении коэффициентов связей между нейронами. В процессе обучения нейронная сеть способна выявлять сложные зависимости между входными данными и выходными, а также выполнять обобщение. Это значит, что в случае успешного обучения сеть сможет вернуть верный результат на основании данных, которые отсутствовали в обучающей выборке.

Для задачи выделения сигналов ФП это будет означать, что в процессе обучения искусственной нейронной сети можно будет подавать сигналы ФП подверженные воздействию различным дестабилизирующим факторам и в случае успешного обучения сеть будет осуществлять выделение сигналов ФП с высокой вероятностью запуска даже при воздействии дестабилизирующих факторов, которые не использовались для обучения. А также в ходе эксплуатации можно дообучить сеть на реальных сигналах.

Интуитивно можно предположить, что чем выше будет частота дискретизации $f_Д$ (больше будет отсчетов на элемент комбинации ФП), тем точнее будут выделяться сигналы ФП и соответственно будет уменьшаться вероятность ложного запуска. Но увеличение количества отсчетов на элемент комбинации ФП будет приводить к увеличению сложности вычисления. Поэтому необходимо использовать однослойную искусственную нейронную сеть, которая состоит из четырех нейронов. Для обучения нейронов нейронной сети должны быть использованы процедуры по ГОСТ 52633.5–2011 [3]. Для разрабатываемого алгоритма это будет означать следующее. Во время обучения на вход нейронов сети будет подаваться несколько примеров образа «Свой» (элементов комбинации ФП) и примеров случайного образа «Чужой».

Обучение должно быть выполнено так, чтобы при поступлении на вход нейрона образа «Свой» он выдавал код «1», а при поступлении на вход нейрона образа «Чужой» выдавал код «0».

Для достижения высокой вероятности выделения ФП комбинация ФП должна быть достаточно длинной, что приводить к значительному увеличению количества входов нейрона даже при низкой частоте дискретизации $f_Д$ и сложности вычисления. Поэтому в разрабатываемом алгоритме, как видно из рис. 1, на вход нейронов будут подаваться отсчеты только двух элементов комбинации ФП. На следующем шаге алгоритма будет выполняться сдвиг комбинации ФП на один элемент и так далее.

Согласно [1] автоматы обучения нейронов должны задавать значения весовых коэффициентов μ_i равному нормированному псевдодискретному входному качеству $Q(v_i)$, соответствующего отсчета элементов входной комбинации:

$$\mu_i = \frac{Q(v_i)}{\sigma_{\text{ЧУЖОЙ}}(v_i)},$$

где $Q(v_i)$ – псевдодискретное входное качество i-го входного отсчета элементов входной комбинации, вычисляемое по формуле:

$$Q(v_i) = \frac{\left|E_{\text{ЧУЖОЙ}}(v_i) - E_{\text{СВОЙ}}(v_i)\right|}{\sigma_{\text{СВОЙ}}(v_i)},$$

где v_i – i-й входной отсчет элементов входной комбинации обучаемого нейрона; $E(\cdot)$ – оператор вычисления математического ожидания; $\sigma(\cdot)$ – оператор вычисления стандартного отклонения.

Математическое ожидание для i-го входного отсчета элементов входной комбинации вычисляется по следующей формуле:

$$E(v_i) = \frac{1}{N} \cdot \sum_{j=1}^{N} v_{ij},$$

где v_{ij} – i-й входной отсчет элементов входной комбинации для j-го образа; N – число образов.

Стандартное отклонение для i-го входного отсчета элементов входной комбинации вычисляется по следующей формуле:

$$\sigma(v_i) = \sqrt{\frac{1}{N} \cdot \sum_{j=1}^{N} \left(v_{ij} - E(v_i)\right)^2}.$$

Исследование алгоритма выделения фазового пуска аппаратуры передачи данных проводилось с использованием программы, написанной на языке MATLAB.

Были проведены исследования работоспособности алгоритма для настроенных нейронов при уровне шума 1,0. На рис. 2–4 приведены распределения выходных значений нейронов при уровнях шума в канале связи 0,1, 0,5 и 1,0 соответственно.

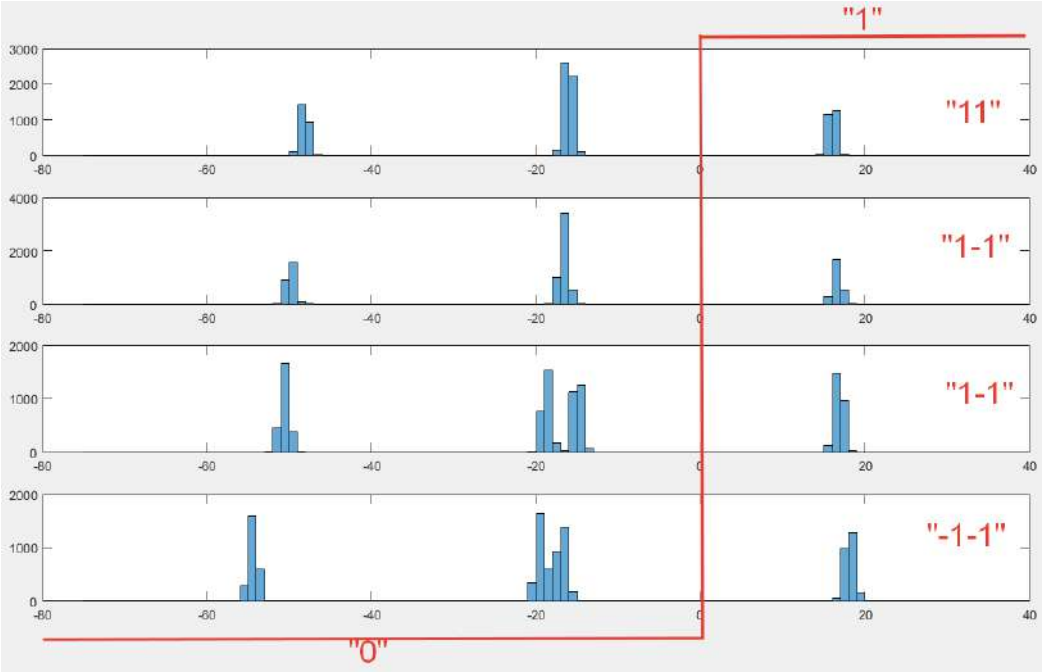


Рис. 2. Распределения выходных значений нейронов при уровне 0,1

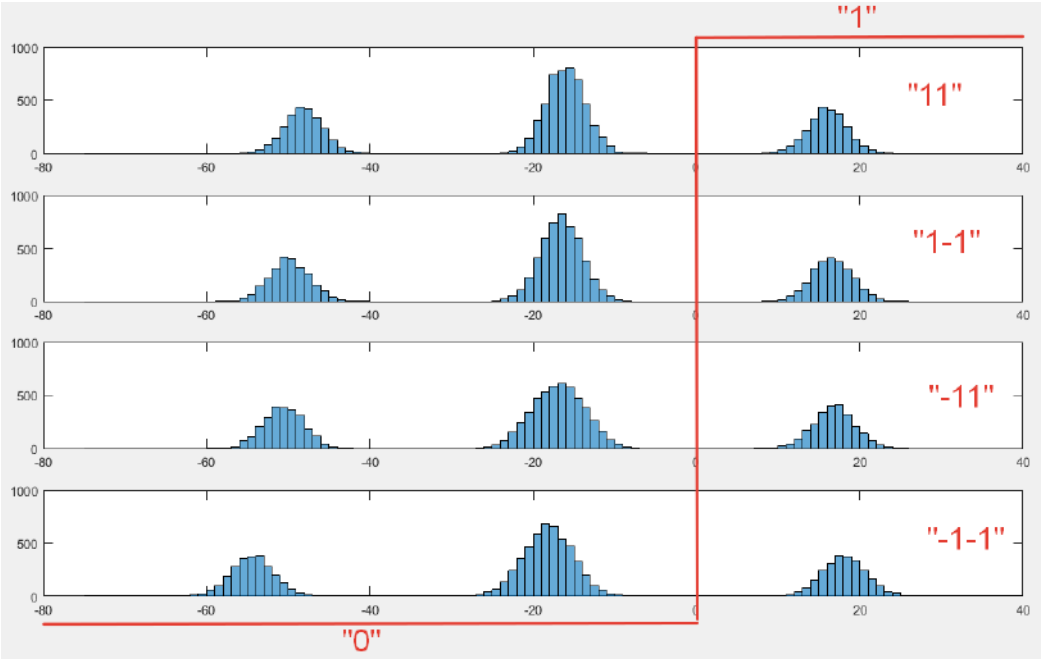


Рис. 3. Распределения выходных значений нейронов при уровне 0,5

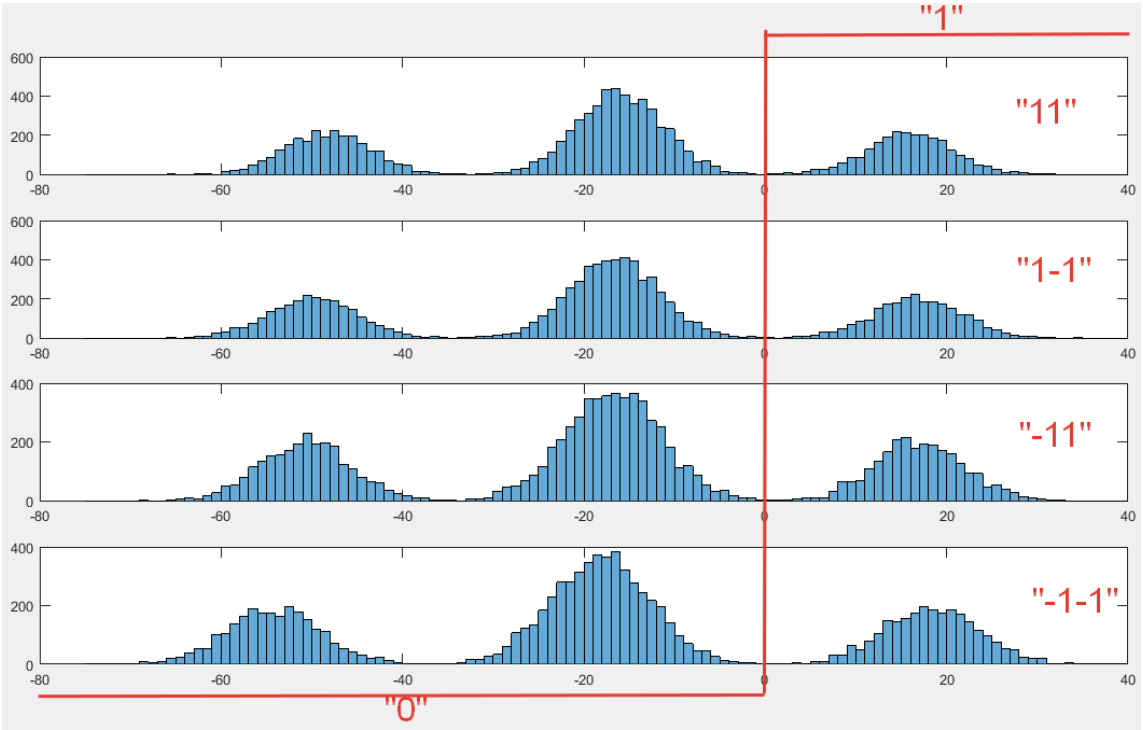


Рис. 4. Распределения выходных значений нейронов при уровне 1,0

Как видно из рисунков с 2 по 4 алгоритм работоспособен, т. е. с использованием искусственной нейронной сети, настроенной при шуме уровнем 1,0 по предложенному алгоритму, можно разделить образы «Свой» и «Чужие» даже при уровне шума 1,0.

Таким образом, используя предложенный алгоритм выделения ФП можно настроить нейроны под текущую помеховую обстановку в канале связи так, чтобы выделение сигналов ФП было всегда верным.

Список литературы

1. Коршунов А. В, Плотников Н. В. Системы телефонной и документальной связи : учеб. пособие. Томск : Изд. Дом ТГУ, 2015. Ч. 1. 118 с.

2. Сердюков П. Н., Бельчиков А. В., Дронов А. Е. [и др.]. Защищенные радиосистемы цифровой передачи информации. М. : АСТ, 2006. 403 с.

3. ГОСТ Р 52633.5–2011. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа. М. : Стандартинформ, 2012. 16 с.

СВЕДЕНИЯ ОБ АВТОРАХ

Астахов Денис Юрьевич, инженер-программист 1 категории, АО «НПП "Рубин"», г. Пенза.

Архипов Александр Владимирович, инженер-программист 2 категории, АО «НПП "Рубин"», г. Пенза.

Бабич Андрей Михайлович, к.т.н., инженер-программист 1 категории, АО «НПП "Рубин"», г. Пенза.

Безяев Александр Викторович, к.т.н., ведущий научный сотрудник, Пензенский филиал АО «НТЦ "Атлас"», г. Пенза.

Борисов Максим Андреевич, инженер-системотехник 1 категории, АО «НПП "Рубин"», г. Пенза.

Бутаев Михаил Матвеевич, д.т.н., профессор, ученый секретарь НТС, АО «НПП "Рубин"», г. Пенза.

Герасин Владислав Юрьевич, ассистент, Финансовый университет при Правительстве Российской Федерации, г. Москва.

Гужова Светлана Андреевна, инженер-программист 2 категории, АО «НПП "Рубин"», г. Пенза.

Демушкин Михаил Олегович, инженер-программист 2 категории, АО «НПП "Рубин"», г. Пенза.

Демушкина Ксения Михайловна, инженер-программист 2 категории, АО «НПП "Рубин"», г. Пенза.

Ерков Михаил Александрович, доцент кафедры общевоенной подготовки Военного учебного центра при Пензенском государственном университете, г. Пенза.

Зверев Олег Владимирович, инженер-программист 3 категории, АО «НПП "Рубин"», г. Пенза.

Зиновьев Дмитрий Михайлович, инженер-электроник 1 категории, АО «НПП "Рубин"», 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.

Зоткина Алёна Александровна, ассистент кафедры программирования, Пензенский государственный технологический университет, г. Пенза.

Иванов Александр Иванович, д.т.н., профессор, научный консультант АО «ПНИЭИ», г. Пенза.

Иванов Алексей Петрович, к.т.н., доцент, заведующий кафедрой технических средств информационной безопасности, Пензенский государственный университет, г. Пенза.

Ильина Ирина Сергеевна, инженер 2 категории, АО «НПП "Рубин"», г. Пенза.

Иниватов Даниил Павлович, ассистент кафедры комплексной защиты информации, Омский государственный технический университет, г. Омск.

Казбаев Антон Валерьевич, начальник отдела, АО «НПП "Рубин"», г. Пенза.

Качалин Сергей Викторович, к.т.н., заместитель начальника отделения, АО «НПП "Рубин"», г. Пенза.

Кирин Александр Сергеевич, студент, Пензенский государственный университет, г. Пенза.

Киселев Антон Михайлович, инженер-программист, АО «НПП "Рубин"», г. Пенза.

Кондаков Сергей Евгеньевич, к.т.н., доцент кафедры ИУ-10, Московский государственный технический университет имени Н. Э. Баумана, г. Москва.

Кочетков Андрей Александрович, начальник сектора, АО «НПП "Рубин"», г. Пенза.

Крутов Алексей Николаевич, доцент кафедры безопасности информационных систем, Самарский национальный исследовательский университет имени академика С. П. Королева, г. Самара.

Кузнецов Вячеслав Ефимович, к.т.н., начальник научно-технического центра, АО «НПП "Рубин"», г. Пенза.

Куликов Сергей Владимирович, научный сотрудник, АО «ПНИЭИ», г. Пенза.

Куц Леонид Валентинович, к.т.н., начальник сектора, АО «НПП «Рубин», г. Пенза.

Майоров Александр Викторович, к.т.н., начальник лаборатории биометрических и нейросетевых технологий, АО «ПНИЭИ», г. Пенза.

Малыгина Елена Александровна, д.т.н., доцент кафедры, Московский государственный технологический университет – МИРЭА, г. Москва.

Малыгин Александр Юрьевич, д.т.н., профессор кафедры радио и спутниковой связи Военного учебного центра при Пензенском государственном университете, г. Пенза.

Масалов Александр Михайлович, студент, Пензенский государственный университет, г. Пенза.

Митрохин Максим Александрович, д.т.н., доцент, заведующий кафедрой вычислительной техники, Пензенский государственный университет, г. Пенза.

Панфилова Ирина Евгеньевна, преподаватель кафедры электронных систем и информационной безопасности, Самарский государственный технический университет, г. Самара.

Папуша Никита Александрович, аспирант, Пензенский государственный университет, г. Пенза.

Пелёвин Сергей Николаевич, инженер-программист, АО «НПП "Рубин"», г. Пенза.

Первушкин Павел Петрович, ведущий инженер-электроник, АО «НПП "Рубин"», г. Пенза.

Романихин Роман Юрьевич, инженер-электроник 1 категории, АО «НПП "Рубин"», г. Пенза.

Савинов Константин Николаевич, старший преподаватель военного учебного центра, Пензенский государственный университет, г. Пенза.

Самарцев Евгений Константинович, студент, Пензенский государственный университет, г. Пенза.

Секретов Максим Викторович, к.т.н., старший научный сотрудник, АО «ПНИЭИ», г. Пенза.

Семёнова Екатерина Николаевна, инженер, АО «НПП "Рубин"», г. Пенза.

Сериков Андрей Васильевич, начальник отделения, АО «НПП "Рубин"», г. Пенза.

Серикова Юлия Игоревна, инженер-программист 1 категории, АО «НПП "Рубин"», г. Пенза.

Серикова Наталья Игоревна, к.т.н., инженер-программист 1 категории, АО «НПП "Рубин"», г. Пенза.

Сергина Ирина Геннадьевна, инженер-программист 1 категории, АО «НПП "Рубин"», г. Пенза.

Строителева Анна Андреевна, соискатель, Пензенский государственный университет, г. Пенза.

Строков Алексей Валерьевич, заместитель начальника управления, ООО «Системы распределенного реестра», г. Москва.

Тарасов Андрей Анатольевич, к.т.н., генеральный директор АО «НПП "Рубин"», г. Пенза.

Тарасов Дмитрий Викторович, к.т.н., доцент кафедры высшей и прикладной математики, Пензенский государственный университет, г. Пенза.

Турыгин Игорь Геннадьевич, к.т.н., начальник сектора, АО «НПП "Рубин"», г. Пенза.

Уткина Мария Юрьевна, инженер-программист 2 категории, АО «НПП "Рубин"», г. Пенза.

Филипов Иван Александрович, преподаватель военного учебного центра, Пензенский государственный университет, г. Пенза.

Хворостухин Сергей Павлович, к.т.н., ведущий научный сотрудник, АО «ПНИЭИ», г. Пенза.

Холопов Денис Андреевич, студент, Пензенский государственный университет, г. Пенза.

Хохлов Игорь Вадимович, инженер 3 категории, АО «НПП "Рубин"», г. Пенза.

Цибин Валерий Евгеньевич, студент, Пензенский государственный университет, г. Пенза.

Чёрный Игорь Валерьевич, к.т.н., главный специалист научно-технического центра, АО «НПП "Рубин"», г. Пенза.

Чудин Кирилл Сергеевич, ассистент кафедры ИУ-10, Московский государственный технический университет имени Н. Э. Баумана, г. Москва.

Шарикова Юлия Валерьевна, инженер-программист 3 категории, АО «НПП "Рубин"», г. Пенза.

СОДЕРЖАНИЕ

Научное издание

Безопасность информационных технологий

Сборник научных статей по материалам
V Всероссийской научно-технической конференции,
посвященной 70-летнему юбилею АО «НПП "Рубин"»

(г. Пенза, 27 сентября 2023 г.)

Том 1

Материалы публикуются в авторской редакции

Корректор *В. В. Чувашова*
Технический редактор *Н. В. Иванов*а
Компьютерная верстка *Н. В. Ивановой*
Дизайн обложки *И. В. Шваревой*

Подписано в печать 27.12.2023.
Формат $60\times84^{1}/_{16}$. Усл. печ. л. 12,67.
Заказ № 698. Тираж 50.

Издательство ПГУ.
440026, г. Пенза, ул. Красная, 40.
Тел.: (8412) 66-60-49, 66-67-77; e-mail: iic@pnzgu.ru