На правах рукописи

Козин Иван Сергеевич

МЕТОД ОБЕСПЕЧЕНИЯ БЕЗОПАСНОСТИ ДАННЫХ ПРИ ИХ ОБРАБОТКЕ В БЛОКЧЕЙН-СИСТЕМЕ ЗА СЧЁТ ПРИМЕНЕНИЯ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

2.3.6. Методы и системы защиты информации, информационная безопасность

Автореферат
диссертации на соискание ученой степени
кандидата технических наук

Санкт-Петербург – 2022

Работа выполнена в Федеральном государственном автономном образовательном учреждении высшего образования «Санкт-Петербургский государственный университет аэрокосмического приборостроения» на кафедре технологий защиты информации.

Научный руководитель: доктор технических наук, доцент
Беззатеев Сергей Валентинович

Официальные оппоненты: **Татарникова Татьяна Михайловна**,
доктор технических наук, профессор,
Санкт-Петербургский государственный
электротехнический университет «ЛЭТИ»
им. В.И. Ульянова (Ленина), кафедра
информационные системы, профессор кафедры

Полтавцева Мария Анатольевна,
доктор технических наук, доцент,
Санкт-Петербургский политехнический университет
Петра Великого, институт кибербезопасности
и защиты информации, доцент

Ведущая организация: Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Сибирский государственный университет
телекоммуникаций и информатики»,
г. Новосибирск

Защита состоится 20 апреля 2022 года в 14.00 на заседании объединенного диссертационного совета 99.2.038.03, созданного на базе Федерального государственного бюджетного образовательного учреждения высшего образования «Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова», Федерального государственного автономного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет аэрокосмического приборостроения», Федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича» по адресу: Санкт-Петербург, пр. Большевиков, д. 22, корп. 1, ауд. 554/1.

С диссертацией можно ознакомиться в библиотеке СПбГУТ по адресу Санкт-Петербург, пр. Большевиков, д. 22, корп. 1 и на сайте www.sut.ru.

Автореферат разослан 18 февраля 2022 года.

Ученый секретарь
диссертационного совета 99.2.038.03,
канд. техн. наук, доцент А.Г. Владыко

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. Наблюдаемое в настоящее время массовое использование средств вычислительной техники в структуре хозяйственного, финансового и экономического управления, а также развитие всемирной электронной коммерции и бизнеса приводят к постоянно возрастающей скорости развития информационных технологий, что способствует непрерывному расширению спектра угроз безопасности информации, обрабатываемой в информационных системах.

В числе наиболее перспективных методов обеспечения хранения, обработки и защиты данных можно выделить методы, учитывающие применение облачных платформах, в т.ч. разработанных с применением технологии «блокчейн».

В частности в России, в целях реализации национальной программы «Цифровая экономика Российской Федерации» и повышения эффективности использования информационно-технологической и коммуникационной инфраструктуры, созданной для предоставления государственных и муниципальных услуг в электронной форме, Правительство Российской Федерации постановило провести в период с 30.08.2019 по 30.12.2020 эксперимент по переводу информационных систем и информационных ресурсов федеральных органов исполнительной власти и государственных внебюджетных фондов в государственную единую облачную платформу.

Необходимость обработки данных в блокчейн-системах предполагает новый эволюционный этап обеспечения безопасности информации и предъявляет к информационным системам дополнительные требования, связанные со спецификой обрабатываемых данных, появлением новых технологий обработки информации, появлением задач, специфических для современного этапа развития информационных технологий, а также с непрерывным расширением спектра потенциальных угроз и, как следствие, устареванием существующих методов защиты данных.

В настоящее время в недостаточной степени изучены модели выявления актуальных угроз нарушения информационной безопасности данных в блокчейн-системе, методы обеспечения безопасности данных в блокчейн-системах, а также методики анализа санкционированного поведения пользователей.

Указанные недостатки существующих моделей, методов и методик защиты данных находятся в противоречии с объективной потребностью обеспечения безопасности данных при их обработке в блокчейн-системе.

Данное противоречие обусловливает существование научной задачи, заключающейся в необходимости разработки методов обеспечения безопасности данных при их обработке в блокчейн-системе. Существующая научная задача обусловила выбор темы данного исследования.

В качестве объекта исследования в настоящей работе выступают процессы обработки и защиты данных в блокчейн-системе, не зависимо от типа данных. Однако в разделе 1 диссертационной работы будут рассмотрены основные требования к разработке и защите информационных систем персональных данных, как в целом соответствующие современному международному уровню достижений науки и техники, а в разделе 3 для демонстрации примера использования предложенного

решения в качестве конкретного типа данных будут рассмотрены персональные данные (далее – ПДн). Актуальность выбора типа данных обусловлена следующим. В мире в целом и в России в частности в качестве одного из основных типов данных, в отношение которых должна обеспечиваться безопасность, можно выделить ПДн. В 2019 году была осуществлена утечка ПДн от 600 млн – до 1,500 млрд субъектов по всему миру (в т.ч. клиентов Facebook, Toyota, FBI, Verification IO LLC), в т.ч. только официально подтверждена утечка свыше 10 млн субъектов в России (в т.ч. клиентов Сбербанка, Альфа-банка, ОТП-банка, ХКФ-банка, ГринМани, Вымпелкома). По данным аналитического агентстве Cybersecurity Ventures, за 2021 года ежегодный ущерб от киберпреступлений составляет примерно 6 млрд долларов США, а количество средств, вкладываемых в кибербезопасность превышает 1 трлн долларов США. В этой связи обеспечение безопасности данных, при их обработке в блокчейн-системах, является задачей ближайшей перспективы, а разработка защищённых блокчейн-систем актуальной задачей.

Степень разработанности темы исследования. Задачам обеспечения безопасности информации при её обработке в информационных системах посвящены многие научные и практические исследования ученых (Беззатеев С.В., Буйневич М.В., Яковлев В.А., Швед В.Г., Крук Е.А., Тюрликов А.М., Матвеев Ю.Н., Заколдаев Д.А., Бирюков В.В., Кармановский Н.С., Мошак Н.Н., Катаржнов А.Д.) и организаций ведущих мировых стран (NIST, Microsoft Corporation, Cisco Systems Inc, АО «НПО «Эшелон», АО «Позитив Текнолоджиз» и т.д.), которые проводятся с 80-х гг. прошлого столетия. Отдельно необходимо отметить высокую степень проработки задач обеспечения безопасности информации с использованием искусственных нейронных сетей в ФГАОУ ВО «Национальный исследовательский университет ИТМО» и, в частности, выделить диссертационные работы, подготовленные под научным руководством Матвеева Ю.Н.

Цели и задачи. Целью диссертационного исследования является обеспечение достоверности данных при их обработке в блокчейн-системе.

Прикладные аспекты поставленной цели связаны:

а) с моделированием угроз:

1) позволяющим определить угрозы, актуальные для блокчейн-систем, а также оказывающие влияние на достоверность данных;

2) в частности в России: позволяющим обеспечить выполнение требований государственных стандартов в части определения актуальных угроз; дополняющим метод определения актуальных угроз, предложенный федеральным органом исполнительной власти, осуществляющим реализацию государственной политики, организацию межведомственной координации и взаимодействия, специальные и контрольные функции в области государственной безопасности по вопросам технической защиты информации (далее – ФОИВ ТЗИ).

б) с обеспечением достоверности данных: при их обработке в блокчейн-системе как на уровне организаций различных форм собственности, так и на уровне государства в целом; в частности, в России при выполнении требований ФОИВ ТЗИ в части контроля ошибочных действий пользователей по вводу данных.

в) с анализом санкционированного поведения пользователей, позволяющим:

1) оперативно выявлять аномалии в поведении пользователей;

2) в частности в России: при обеспечении выполнения требований ФОИВ ТЗИ в части анализа зарегистрированных событий безопасности и реагирования на них; при защите от угроз, представленных в банке данных угроз ФОИВ ТЗИ, связанных с подменой доверенного пользователя и его действий путём обмана.

Диссертационная работа является продолжением и развитием результатов исследований научных коллективов в этой области, проводимых в течение последних 10-ти лет.

Научная новизна. Представленный метод обеспечения безопасности данных при их обработке в блокчейн-системе за счёт применения искусственных нейронных сетей включает в себя следующие материалы, составляющие научную новизну:

1) представление угроз, отличное от классического (предложенного ФОИВ ТЗИ) учётом: специфических угроз данным, обрабатываемым в блокчейн-системе; числовых значений степени опасности нарушения отдельных характеристик безопасности (конфиденциальности, целостности, доступности, достоверности);

2) концептуально новым подходом к достижению консенсуса, включающим процедуру автоматизированной оценки рисков внесения и обработки недостоверных данных, разработанную с использованием теории искусственных нейронных сетей и теории нечетких множеств;

3) уникальным составом формализованных характеристик поведения пользователя (составом входных характеристик нейронной сети) и параметрами нейронной сети.

Теоретическая и практическая значимость.

Теоретическая значимость диссертационной работы состоит в следующем:

1. Установлена и формализована зависимость между: угрозами, актуальными для данных, обрабатываемых в блокчейн-системе; ущербом от потенциальных угроз блокчейн-системе; степенью опасности нарушения отдельных характеристик безопасности (в т.ч. достоверности); составом деструктивных воздействий; степенью важности данных.

2. Расширен класс методов обеспечения достоверности данных в части выявления недостоверных персональных данных при их вводе в блокчейн-систему за счёт искусственной нейронной сети.

3. Формализовано поведение пользователя и доказана возможность выявления аномалий в поведении пользователя при помощи искусственных нейронных сетей.

Практическая значимость диссертационной работы состоит в том, что ее результаты позволяют:

а) определить угрозы, актуальные для блокчейн-систем, а также оказывающие влияние на достоверность данных;

б) обеспечить достоверность данных при их обработке в блокчейн-системе как на уровне организаций различных форм собственности, так и на уровне государства в целом;

в) оперативно выявлять аномалии в поведении пользователей;

г) в частности в России:

1) обеспечить выполнение требований государственных стандартов в части определения актуальных угроз;

2) выполнить требования ФОИВ ТЗИ: по контролю ошибочных действий пользователей при вводе персональных данных (в части анализа зарегистрированных событий безопасности и реагирования на них); при защите от угроз, представленных в банке данных угроз ФОИВ ТЗИ, связанных с подменой доверенного пользователя и его действий путём обмана;

3) дополнить метод определения актуальных угроз, предложенный ФОИВ ТЗИ.

Результаты научных исследований были использованы при проведении научно-исследовательской и опытно-конструкторских работ на предприятиях ПАО «Интелтех», ООО «СИГМА». Также результаты диссертационной работы использованы в учебном процессе ФГАОУ ВО «СПбГУАП».

Методология и методы исследования. В процессе исследования использовались методы теории нечётких множеств, теории искусственных нейронных сетей и вычислительной математики.

Положения, выносимые на защиту.

1. Модель выявления актуальных угроз нарушения информационной безопасности данных, обрабатываемых в блокчейн-системе.
2. Метод обеспечения достоверности персональных данных, обрабатываемых в блокчейн-системе.
3. Методика анализа санкционированного поведения пользователей информационной системы.

Степень достоверности и апробация результатов. Обоснованность и *достоверность* выносимых на защиту полученных новых положений, выводов и рекомендаций научного и практического характера обусловлены и подтверждаются корректностью математического обоснования проведенных исследований и системным подходом к решению поставленных задач, в том числе: теоретически обоснованным выбором основных методов обеспечения защищённости информации и созданием на их основе модели выявления актуальных угроз, метода определения достоверности данных и методики анализа санкционированного поведения пользователей; применением обоснованного математического аппарата, уточненного с учетом специфики решения указанных задач; математическими доказательствами и результатами экспериментальной проверки работоспособности предложенного метода анализа поведения пользователей на реальных исходных данных; всесторонним анализом работ и согласованностью с известными решениями в данной предметной области; обсуждением в открытой печати, апробацией на всероссийской конференции.

Апробация результатов. Основные результаты проведённых исследований обсуждались на конференциях: Российская научно-техническая конференция «Новые информационные технологии в системах связи и управления» (г. Калуга, 2016); четвёртая конференция по программной инженерии и организации информации «Software Engineering and Information Management, SEIM-2019» (г. Санкт-Петербург, 2019).

Публикации. Основные результаты диссертационной работы опубликованы в 11 печатных трудах, из них 5 статей опубликовано в рецензируемых научных изданиях, рекомендованных ВАК, 5 в других изданиях и материалах конференций и 1 отчет о НИР.

Диссертация соответствует пунктам 1–3, 5, 7, 9, 13, 14 паспорта специальности «Методы и системы защиты информации, информационная безопасность».

Личное участие соискателя Основные научные результаты диссертационного исследования получены автором лично.

Структура и объем диссертации. Диссертация состоит из введения, 5 разделов, заключения, списка литературы и 4 приложений. Материал изложен на 147 страницах, включает 21 таблицу и 27 рисунков. Список литературы содержит 131 наименование.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность диссертационной работы, описано состояние исследуемого противоречия, сформулированы цели и задачи работы, определены научная новизна и теоретическая значимость результатов, описана область их применения, представлены основные положения, выносимые на защиту, приведены сведения об апробации работы и о публикациях по теме работы, указаны пункты паспорта специальности, которым соответствует работа, а также определена степень личного участия автора.

В первой главе: 1. Представлено обобщённое описание требований к защите конфиденциальных данных на примере персональных данных при их обработке в информационных системах, а также методов разработки защищённых информационных систем персональных данных. 2. В качестве одной из ключевых тенденций развития информационных систем определена необходимость перевода информационных систем на облачную платформу. 3. На основании представленного обобщённого описания требований и методов к обеспечению безопасности данных, а также с учётом намеченных тенденций развития информационных систем, обоснована актуальность разработки метода обеспечения безопасности данных при их обработке в блокчейн-системе за счёт применения искусственных нейронных сетей, учитывающего необходимость совершенствования существующего метода определения актуальных угроз, предусматривающего обеспечение достоверности данных при их обработке в блокчейн-системе, а также предполагающего обеспечение анализа санкционированного поведения пользователей информационной системы. 4. Утверждается, что разработка метода обеспечения безопасности данных при их обработке в блокчейн-системе за счёт применения искусственных нейронных сетей, учитывающего специфику блокчейн-систем при разработке модели угроз, предусматривающего обеспечение достоверности данных, а также проведение анализа санкционированного поведения пользователей, позволит обеспечить безопасность данных при их обработке в блокчейн-системе. 5. Сформулирована цель диссертационной работы. Представлена оценка предложенных автором решений,

позволяющая сделать вывод о достижении поставленной цели. Выделены задачи, за решение которых берётся автор и решение которых позволит заложить научные и практические основы обеспечения безопасности данных при их обработке в блокчейн-системе, что в итоге имеет важное значение для обеспечения безопасности данных вне зависимости от государства, в котором предполагается эксплуатация блокчейн-системы.

Во второй главе установлена и формализована зависимость между: угрозами, актуальными для данных, обрабатываемых в блокчейн-системе; составом деструктивных воздействий; степенью важности данных; степенью опасности нарушения отдельных характеристик безопасности (в т.ч. достоверности); ущербом от потенциальных угроз блокчейн-системе.

В состав основных угроз блокчейн-системе предложено включить следующие угрозы: атака 51%; двойная трата; атака Сибиллы; DDoS; взлом криптографических алгоритмов; атака за счёт маршрутизации; внесение в реестр некорректной информации; угрозы за счёт уязвимостей в смарт-контрактах (в т.ч. связанные: с генерацией ключей и прекращением выполнения кода; с вызовами функций другого смарт-котракта и пр.).

Предлагается выделить следующие деструктивные воздействия, способные повлиять на безопасность блокчейн-системы: ознакомление; частичное блокирование (снижение показателей скорости работы); полное блокирование (прекращение работы); модификация; удаление.

В таблице 1 автором предложена градация степеней важности объекта защиты в зависимости от возможных последствий, к которым может привести нарушение характеристик безопасности объекта защиты. Определение степени важности объекта защиты необходимо для определения опасности нарушения характеристик безопасности объектов защиты.

Таблица 1 – Сопоставление последствий от нарушения безопасности данных и степени важности объект защиты

Последствия, к которым может привести нарушение безопасности данных	Вывод о степени важности объекта
Высокий ущерб	1-я степень
Значительный ущерб	2-я степень
Умеренный ущерб	3-я степень
Незначительный ущерб	4-я степень

В качестве математического аппарата для присвоения числовых значений характеристикам опасности нарушения характеристик безопасности объектов защиты предлагается применить теорию нечётких множеств.

Примерные значения опасности нарушения характеристик безопасности, предлагаемые автором на основании применения теории нечётких множеств, представлены в таблице 2 и, для наглядности, на диаграмме Заде[1] (рисунок 1).

[1] Диаграмма Заде – форма представления нечёткого множества в виде графика его функции принадлежности в координатах (u, $x(u)$) на плоскости.

Для уточнения примерных числовых значений функций принадлежности[2] множествам опасности были сделаны следующие допущения: на конфиденциальность не могут повлиять блокирование, модификация и удаление объекта защиты; на целостность не могут повлиять ознакомление и блокирование объекта защиты; на доступность не может повлиять ознакомление с объектом защиты; при полном блокировании опасность нарушения доступности выше, чем при частичном; при модификации опасность нарушения выше, чем опасность нарушения доступности.

Таблица 2 – Примерные значения опасности нарушения характеристик безопасности

Функции принадлежности множествам опасности	Примерные значения опасности нарушения характеристик безопасности			
	Нулевая опасность	Низкая опасность	Средняя опасность	Высокая опасность
$\mu_A(x)$	0,20	0,40	0,70	0,90
$\mu_B(x)$	0,40	0,70	0,90	0,70
$\mu_C(x)$	0,70	0,90	0,70	0,40
$\mu_D(x)$	0,90	0,70	0,40	0,20

При уточнении значений функций принадлежности множествам опасности был осуществлён сдвиг по оси x на 0,1–0,2 (для $K_{do\gamma}$ в случае частичного блокирования и модификации). Таким образом, с учётом значений, представленных в таблице 2, уточнённые числовые значения опасности нарушения характеристик безопасности, выраженные максимальными значениями соответствующих функций принадлежности, приняли значения, представленные в таблице 3 (см. ниже).

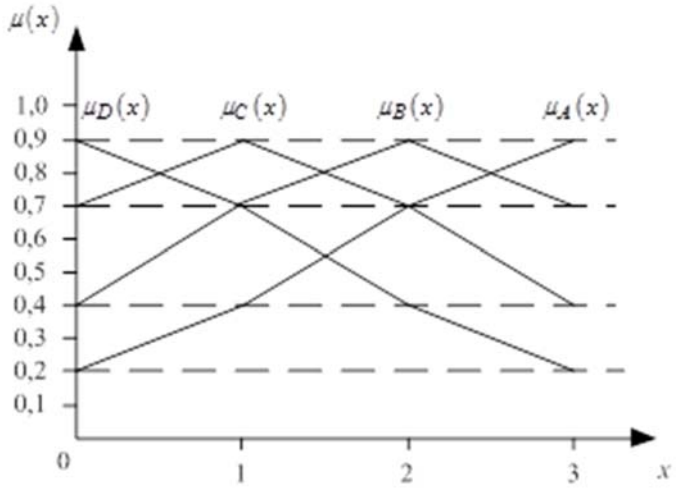

Рисунок 1 – Примерные значения опасности нарушения характеристик безопасности

Таким образом:

1. Общая опасность d-го деструктивного воздействия выражена коэффициентом K_{do}. Коэффициент общего деструктивного воздействия K_{do} в соответствии с формулой, предложенной органом государственной власти, уполномоченным по вопросам технической защиты информации, определяется следующим соотношением:

$$K_{do} = \theta_{d\alpha} K_{do\alpha} + \theta_{d\beta} K_{do\beta} + \theta_{d\gamma} K_{do\gamma} + \theta_{d\varepsilon} K_{do\varepsilon}, \quad (1)$$

где $\theta_{d\alpha}$, $\theta_{d\beta}$, $\theta_{d\gamma}$, $\theta_{d\varepsilon}$ – функции, значения которых равны единице, если в результате реализации d-го деструктивного воздействия нарушается конфиденциальность,

[2] В рамках теории нечётких множеств, характеристические функции принадлежности отражают субъективный взгляд специалиста на решаемую задачу.

целостность, доступность или достоверность объекта соответственно; $K_{do\alpha}$, $K_{do\beta}$, $K_{do\gamma}$, $K_{do\varepsilon}$ – коэффициенты деструктивного воздействия по отношению к конкретным характеристикам безопасности.

Таблица 3 – Уточнённые числовые значения опасности нарушения характеристик безопасности

Деструктивные воздействия	Степень важности объекта защиты	Числовые значения характеристик безопасности			
		Конфиденциальность ($K_{do\alpha}$)	Целостность ($K_{do\beta}$)	Доступность ($K_{do\gamma}$)	Достоверность ($K_{do\varepsilon}$)
Ознакомление	Первая	0,9	–	–	–
	Вторая	0,7	–	–	–
	Третья	0,5	–	–	–
	Четвёртая	0,3	–	–	–
Блокирование частичное	Первая	–	–	0,7	–
	Вторая	–	–	0,5	–
	Третья	–	–	0,3	–
	Четвёртая	–	–	0,1	–
Блокирование полное	Первая	–	–	0,9	–
	Вторая	–	–	0,7	–
	Третья	–	–	0,5	–
	Четвёртая	–	–	0,3	–
Модификация	Первая	–	0,9	0,8	0,9
	Вторая	–	0,7	0,6	0,7
	Третья	–	0,5	0,3	0,5
	Четвёртая	–	0,3	0,1	0,3
Удаление	Первая	–	0,9	0,9	–
	Вторая	–	0,7	0,7	–
	Третья	–	0,5	0,5	–
	Четвёртая	–	0,3	0,3	–

2. Ущерб от реализации угрозы t по отношению к объекту o выражен коэффициентом ущерба D_{to}, определяемым совокупностью коэффициентов K_{do} деструктивных воздействий (d_1, d_2, d_3, d_4, d_5), к которым может привести реализация угрозы. Значение коэффициента ущерба D_{to} от реализации угрозы определяется следующим соотношением (в соответствии с формулой, предложенной органом государственной власти, уполномоченным по вопросам технической защиты информации):

$$D_{to} = K_{d_1 o} + K_{d_2 o} + K_{d_3 o} + K_{d_4 o} + K_{d_5 o}. \quad (2)$$

3. Опасность угрозы t выражается отношением ущерба D_{to}, к которому может привести её реализация, к приемлемому ущербу D_a:

$$K_{to} = \frac{D_{to}}{D_a}. \quad (3)$$

При этом предполагается, что приемлемый ущерб D_a не может принимать значение, равное 0, а опасность угрозы не может принимать значение равное 1.

4. Для определения актуальности угрозы числовые характеристики опасности необходимо интерпретировать вербально. Для обеспечения вербальной интерпретации числовых значений автором предлагается применить теорию нечётких множеств. Общая форма записи нечётких подмножеств будет иметь следующий вид:

$$\begin{aligned} A &= \sum_{x=0}^{1} \mu_A(x)/x = \sum_{x=0}^{0} 0/x + \sum_{x=0,33}^{0,33} 0,2/x + \sum_{x=0,66}^{0,66} 0,7/x + \sum_{x=1}^{1} 0,9/x \\ B &= \sum_{x=0}^{1} \mu_B(x)/x = \sum_{x=0}^{0} 0,2/x + \sum_{x=0,33}^{0,33} 0,7/x + \sum_{x=0,66}^{0,66} 0,9/x + \sum_{x=1}^{1} 0,2/x \\ C &= \sum_{x=0}^{1} \mu_C(x)/x = \sum_{x=0}^{0} 0,7/x + \sum_{x=0,33}^{0,33} 0,9/x + \sum_{x=0,66}^{0,66} 0,7/x + \sum_{x=1}^{1} 0,2/x \end{aligned} \quad (4)$$

где A – множество высокой опасности, B – множество средней опасности, C – множество низкой опасности, $\mu_A(x)$ – характеристическая функция принадлежности множеству высокой опасности, $\mu_B(x)$ – характеристическая функция принадлежности множеству средне опасности, $\mu_C(x)$ – характеристическая функция принадлежности множеству низкой опасности.

Полученные автором показатели интерпретации опасности угрозы представлены в таблице 4.

Таблица 4 – Вербальная интерпретация опасности угрозы

Интервал показателя опасности угрозы	Вербальная интерпретация показателя
0,67 – 0,99	Высокая
0,34 – 0,66	Средняя
0,01 – 0,33	Низкая

Определение актуальности угрозы должно осуществляться в порядке, определённом ФОИВ ТЗИ – с учётом описания возможных сценариев.

В третьей главе расширен класс методов обеспечения достоверности данных в части выявления недостоверных ПДн при их вводе в блокчейн-систему за счёт применения искусственной нейронной сети.

Метод включает предложения по построению общей архитектуры распределенного реестра персональных данных (далее – РРПДн), порядку хранения данных, способу достижения консенсуса, обобщённому порядку внедрения и развития системы, а также расчёту вероятности компрометации ПДн.

Предполагается, что: ПДн должны обрабатываться в блокчейн-системе не менее, чем в течение всей жизни субъекта ПДн; на протяжении всего времени обработки ПДн должна обеспечиваться их безопасность (конфиденциальность, доступность, целостность и достоверность).

Построение защищённой РРПДн сводится к решению следующих задач: определение состава ПДн, обработку которых целесообразно осуществлять в РРПДн; определение общей архитектуры РРПДн; определение порядка хранения данных; определение механизма достижения консенсуса (в т.ч. порядка вознаграждения пользователей, обеспечивающих работу РРПДн, а также порядка автоматизированной оценки рисков обработки недостоверных ПДн); выбор способа вычисления хеш-функции; определение общего порядка развития РРПДн.

В качестве примера предлагается рассмотреть РРПДн, предназначенную для обеспечения оперативного взаимодействия между физическими лицами, организациями и предприятиями, в т.ч.: 1) хранения ПДн граждан РФ; 2) оперативного доступа граждан РФ к своим ПДн; 3) оперативного предоставления гражданами РФ доступа к своим ПДн с целью: обеспечения идентификации личности (например, при приобретении товаров и услуг); обеспечения доступа к сведениям об образовании и профессиональных навыках (например, при трудоустройстве, поступлении на учёбу); создания и обмена умными активами; создания и выполнения умных контрактов (например, при приобретении или предоставлении различных услуг).

Для достижения указанной цели в РРПДн необходимо обрабатывать следующие сведения: содержащиеся в паспорте гражданина РФ (фамилия, имя, отчество, пол, дата и место рождения, серия и номер документа, удостоверяющего личность, а также об органе его выдавшем, сведения о месте регистрации); об образовании (наименование учебного заведения, факультета и специальности, год окончания, состав изученных дисциплин и общая успеваемость, дополнительная информация (сведения о внеучебной деятельности, участии в конкурсах и т.п.); о профессиональных навыках (наименование мест работы, подразделений и должностей, годы работы, состав должностных обязанностей, ключевые навыки); о финансовом состоянии (сведения о находящихся во владении объектах недвижимости, транспортных средствах и т.п.).

Таким образом, в РРПДн предполагается обработка «специальных», «биометрических», «общедоступных» и «иных» ПДн.

С учётом цели и состава ПДн, а также больших объёмов данных, которые предполагается обрабатывать в РРПДн, предлагается:

1) выделить в РРПДн несколько самостоятельных частных (не публичных) цепочек блоков, по одной для каждой предметной области: мастерчейн ЦБ-И для идентификационных данных (категории ПДн «биометрические» и «иные»); воркчейн ЦБ-О для данных об образовании (категории ПДн – «общедоступные», «иные»); воркчейн ЦБ-П для данных о профессиональных навыках (категории ПДн – «специальные», «общедоступные», «иные»); воркчейн ЦБ-А для данных об активах (категории ПДн – «иные»); воркчейн ЦБ-К для данных об умных контрактах (категории ПДн – «иные»);

2) в качестве основы мастерчейна использовать технологию блокчейн второго поколения, т.к. использование блоков вместо транзакций позволяет уменьшить объём трафика и нагрузку на вычислительные ресурсы узлов сети;

3) в качестве основы воркчейнов ЦБ-О, ЦБ-П, КБ-А и ЦБ-К использовать технологию блокчейн третьего поколения с применением прямых ациклических

графов, т.к. такой метод позволит содержать в блоках цепочек части файлов и ссылки на другие блоки.

В состав каждого блока предлагается включить 200 записей.

С учётом статистических данных о населении России за период 2018–2020 годы для агломерации в 1 000 000 человек предполагается следующее:

а) каждые день необходимо будет вносить следующие изменения в состав субъектов ПДн: добавлять новорождённых – 35 человек; изменять статус на «умер» – 38 человек; изменять сведения о месте проживания – 75 человек (38 выехали, 37 въехали).

б) в состав обучающихся будут входить: 110 093 учеников учреждений среднего общего образования; 20 514 студентов учреждений среднего специального образования; 27 735 студентов учреждений высшего образования; 41 198 студентов учреждений дополнительного профессионального образования (в т.ч. осуществляющих повышение квалификации и проходящих профессиональную переподготовку); 621 аспирант.

в) работающими людьми будут являться 563 000 человек.

На основании вышеизложенного в таблице 5 представлены предложения по количеству блоков, которые представляется целесообразным сформировать изначально (при создании РРПДн), а также предложения по частоте создания новых блоков.

Таблица 5 – Предложения по количеству и частоте создания новых блоков

Наименование цепочки блоков	Объём одного блока, кБ	Количество изначально созданных блоков	Примерный объём цепочки блоков, ГБ		Средняя примерная частота создания новых блоков
			на старте	увеличение в год	
ЦБ-И	300	6 250	1,9	0,1	1 раз в день
ЦБ-О	2 400	1 001	2,4	2,6	3 раза в день
ЦБ-П	2 400	2 815	6,8	7,0	8 раз в день
ЦБ-К	200	50 000	10,0	7,6	1 раз в 10 минут
ЦБ-А	200	50 000	10,0	7,6	1 раз в 10 минут

Таким образом, примерный объем актуальных данных РРПДн составит до 31,1 ГБ при вводе в эксплуатацию и до 24,9 ГБ ежегодного прироста.

Создание новых блоков не должно являться задачей, требующей значительных вычислительных ресурсов (как в случае применения метода подтверждения работы (Proof of Work)). С учётом специфики рассматриваемой блокчейн-системы наиболее подходящим представляется алгоритм Proof of Authority, предназначенный для обеспечения работы частных сетей и позволяющий выделять привилегированных валидаторов. Предлагается расширить его функциональные возможности процедурой автоматизированной оценки достоверности вносимых в блокчейн-систему данных.

В рамках решаемой задачи предлагается рассмотреть характеристики x_n как: характеристические функции принадлежности $\mu_A(u)$ множеству значений A, сигнализирующих о повышенной вероятности ввода недостоверных ПДн, заданные на универсальном множестве U; и принимающие значения, равные единице, на тех

элементах множества U, которые принадлежат множеству A, и значения, равные нулю, на тех элементах, которые не принадлежат множеству A:

$$\mu_A(u) = \begin{cases} 1, \textit{ если } u \in A \\ 0, \textit{ если } u \notin A \end{cases}.$$

При этом для каждой функции принадлежности должны рассматриваться свои множества. Далее в качестве примеров будут рассмотрены четыре характеристические функции принадлежности: $\mu_{A_a}(u)$ – функция принадлежности множеству значений степени связи узла консенсуса и объекта подтверждения, при которой складываются наиболее благоприятные условия для вступления в сговор; $\mu_{A_b}(u)$ – функция принадлежности множеству значений промежуточных подтверждений, сигнализирующих о повышенной вероятности участия в сговоре; $\mu_{A_c}(u)$ – функция принадлежности множеству значений вознаграждений для объекта подтверждения, оказывающих наибольшую мотивацию для вступления в сговор; $\mu_{A_d}(u)$ – функция принадлежности множеству значений суммарной надёжности узла консенсуса и объекта подтверждения, создающей минимальные предпосылки к вступлению в сговор.

На рисунке 2 представлена диаграмма Заде, демонстрирующая возможную зависимость значений характеристической функции принадлежности множеству значений степени связи узла консенсуса и объекта подтверждения, при которой складываются наиболее благоприятные условия для вступления в сговор, от степени связи узла консенсуса и объекта подтверждения, на которой: U_a – множество значений степени связи узла консенсуса и объекта подтверждения, $U_a = [u_a, u_a \in R: 0 \le u_a \le 10]$; A_a – множество значений степени связи узла консенсуса и объекта подтверждения, при которой складываются наиболее благоприятные условия для вступления в сговор; $\mu_{A_a}(u_a)$– характеристическая функция принадлежности множеству значений степени связи узла консенсуса и объекта подтверждения, при которой складываются наиболее благоприятные условия для вступления в сговор; $x(u_a)$ – значения характеристической функции принадлежности $\mu_{Aa}(u_a)$ множеству значений степени связи узла консенсуса и объекта подтверждения, при которой складываются наиболее благоприятные условия для вступления в сговор.

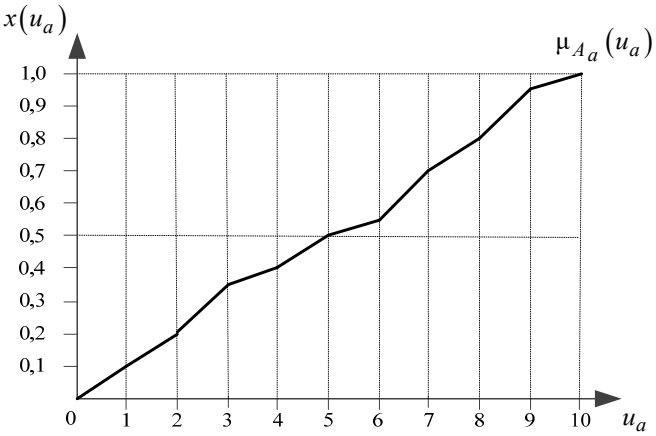

Рисунок 2 – Зависимость степени связи объекта подтверждения и узла консенсуса с благоприятными условиями вступления в сговор

В предложенном примере предполагается, что различным степеням связи соответствуют различные числовые значения u_a, представленные на оси абсцисс:

отсутствие связи – 0; наличие одного общего знакомого – 1; наличие от двух до пяти общих знакомых – 2; наличие более чем шести общих знакомых – 3; не близкие родственники – 4; совместная учёба в школе – 5; совместная учёба в ВУЗе – 6; совместная работа – 7; совместное прохождение военной службы – 8; общий круг интересов – 9; близкие родственники – 10. Конкретные числовые значения взяты в качестве примера и предназначены для демонстрации общего принципа формирования характеристической функции принадлежности и будущих входных значений нейронной сети.

Поскольку с усилением степени связи узла консенсуса и объекта подтверждения (с возрастанием значения u_a) складываются наиболее благоприятные условия для вступления в сговор (возрастает значение $x(u_a)$), функция принадлежности является возрастающей.

Для более удобной интерпретации данных при формировании обучающей выборки, и, при необходимости, её нормирования, представляется целесообразным привести полученные результаты к общей форме записи нечётких подмножеств.

Таким образом, общая форма записи нечетких подмножеств будет иметь следующий вид:

$$A_a = \sum_{u_a=0}^{10} \mu_{A_a}(u_a)/u_a = \sum_{u_a=0}^{0} 0,00/u_a + \sum_{u_a=1}^{1} 0,20/u_a + \sum_{u_a=2}^{2} 0,20/u_a + \sum_{u_a=3}^{3} 0,35/u_a + \sum_{u_a=4}^{4} 0,40/u_a + \\ + \sum_{u_a=5}^{5} 0,50/u_a + \sum_{u_a=6}^{6} 0,55/u_a + \sum_{u_a=7}^{7} 0,70/u_a + \sum_{u_a=8}^{8} 0,75/u_a + \sum_{u_a=9}^{9} 0,95/u_a + \sum_{u_a=10}^{10} 1,00/u_a \quad (5)$$

где: запись вида $\sum_{u_a=0}^{10} \mu_{Aa}(u_a)/u_a$ не предполагает сумму, но предполагает объединение по всем элементам конечного несущего множества значений u_a; в верхней и нижней частях знака суммы $\sum_{u_a=2}^{2} 0,20/u_a$ указывается значение точки u_a на универсальном множестве U; справа от знака суммы $\sum_{u_a=2}^{2} 0,20/u_a$ до знака / указывается значение характеристической функции принадлежности μ_{A_a} в точке u_a.

Числовые значения функции принадлежности и точек u в дальнейшем будут использоваться при формировании обучающих выборок.

При формировании обучающих выборок представляется целесообразным определить значения, которые могут негативно сказаться на процессе обучения нейронной сети – значения, не позволяющие сделать однозначный вывод об аномальности поведения пользователя. В теории нечётких множеств такие значения определяются точками перехода[3] ($x(u_a) = 0,5$). Для функции принадлежности $\mu_{A_a}(u_a)$ такой точкой является $u_a = 5$.

Аналогичным образом должны быть определены числовые значения по другим входящим сигналам.

---

[3] Точкой перехода нечёткого множества A называют элемент множества U, на котором $\mu_A(u) = 0,5$.

В четвёртой главе предложена актуальная методика анализа санкционированного поведения пользователей информационной системы, основанная на применении технологии машинного обучения (теория искусственных нейронных сетей).

Предполагается, что пользователь информационной системы (оператор РРПДн, являющейся представителем организации, создающей новые блоки и добавляющей их в цепочку) обладает набором характеристик, совокупность которых выражает его уникальное типовое поведение. К таким характеристикам предлагается отнести: набор данных, с которыми работает пользователь (файлы, папки, сетевые объекты, интернет-сайты и т.п.); место осуществления доступа к информационной системе персональных данных (конкретный компьютер, № помещения, здание, город, страна и т.п.); набор действий, которые выполняет пользователь (чтение, запись, копирование, модификация и т.п.); время, в которое осуществляется доступ или выполняются определенные действия (время суток, день недели, определенные числа и т.п.); общую продолжительность выполняемых в течение определенного времени действий.

Предложенный набор характеристик не является исчерпывающим, но позволяет построить уникальную модель поведения пользователя.

Отступление от модели поведения (выявление аномалий в поведении) может свидетельствовать о совершении противоправных действий. Примерами таких действий являются: массовое удаление материалов, к которым имеет доступ пользователь (практикуется многими недовольными работниками при увольнении); использование чужой учетной записи (практикуется пользователями, несерьезно относящимися к правилам разграничения доступа); беспорядочное ознакомление или копирование корпоративной информации (практикуется любопытными пользователями и инсайдерами).

Каждая из представленных характеристик пользователя может быть рассмотрена применительно к группе пользователей. Такой метод может найти применение при сговоре среди пользователей и совершении санкционированных неправомерных действий, распределенных среди нескольких человек и потому особо затруднительных в выявлении.

Каждую характеристику пользователя (или группы пользователей) можно описать в виде коэффициентов x_n, $n \in \{1; 5\}$, где n выражает порядковый номер характеристики: x_1 — набор данных, с которыми работает пользователь; x_2 — точка доступа пользователя к системе; x_3 — набор совершаемых пользователем действий; x_4 — время осуществления доступа; x_5 — общая продолжительность проводимых работ.

Необходимо: определить тип необходимой нейросети; разработать метод присвоения числовых значений входным сигналам ИНС (x_1 – x_5), отражающим поведение пользователя или группы пользователей; определить необходимое количество слоёв нейросети (r); определить необходимое количество нейронов в слоях нейросети ($m_1, \ldots, m_r$); выбрать метод обучения нейросети; выбрать виды необходимых активационных функций; выбрать область значений выходного сигнала NET,

сигнализирующего о наличии аномалий в поведении пользователя или группы пользователей.

Поскольку на сегодняшний день не существует строгой теории по выбору ИНС, за основу разрабатываемой ИНС предлагается взять хорошо изученный многослойный полносвязный персептрон без обратных связей.

Для присвоения числовых значений входным сигналам ИНС предлагается использовать математический аппарат теории нечетких множеств, позволяющий присваивать вербальным характеристикам (более свойственно, менее свойственно и т.п.) числовые значения.

Далее в качестве примеров будут рассмотрены четыре характеристические функции принадлежности: $\mu_{A_e}(u)$ – функция принадлежности множеству значений времени суток, в которое доступ к ресурсам информационной системы является для пользователя аномальным; $\mu_{A_f}(u)$ – функция принадлежности множеству данных, работа с которыми является для пользователя аномальной; $\mu_{A_g}(u)$ – функция принадлежности множеству значений, символизирующих рабочие места, доступ с которых является для пользователя аномальным; $\mu_{A_h}(u)$ – функция принадлежности множеству действий, которые являются для пользователя аномальными; $\mu_{A_i}(u)$ – функция принадлежности множеству значений продолжительности работы пользователя, которая является для пользователя аномальной.

На рисунке 3 представлена диаграмма Заде, демонстрирующая возможную зависимость значений характеристической функции принадлежности множеству значений аномального поведения от времени доступа пользователя к ресурсам сети, на которой: U_e – множество значений времени суток, в которое может быть осуществлен доступ к ресурсам сети, $U_e = \{u_e, u_e \in R: 0 \le u_e \le 24\}$; A_e – множество значений времени суток, доступ в которое аномален для конкретного пользователя; $\mu_{Ae}(u_e)$ – характеристическая функция принадлежности множеству значений времени суток, доступ в которое аномален для конкретного пользователя, $\mu_{Ae}(u_e) \in \{0;1\}$; $x(u_e)$ – значения характеристической функции принадлежности $\mu_{Ae}(u_e)$ множеству значений времени суток A_e, доступ в которое аномален для конкретного пользователя.

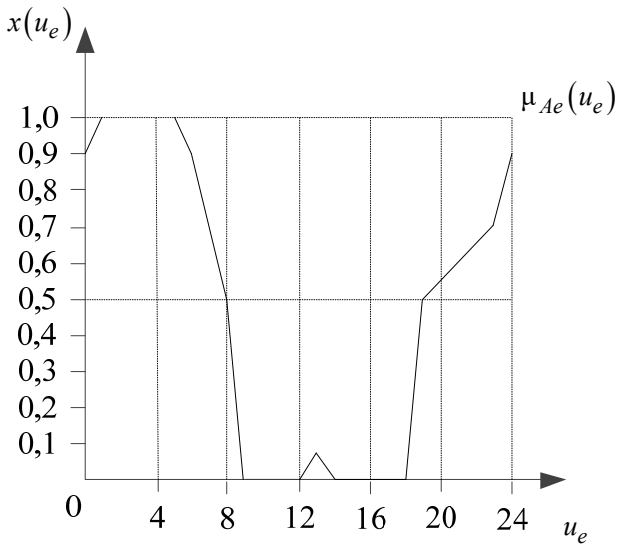

Рисунок 3 – Зависимость между временем доступа и аномальностью поведения

В предложенном примере предполагается следующее: продолжительность рабочего дня пользователя, представленная на оси абсцисс, составляет девять часов (с девяти утра до шести вечера), включая перерыв на обед (в районе часа дня);

пользователю не свойственно приходить на работу раньше начала рабочего дня, задерживаться по окончании рабочего дня и работать в около обеденное время.

Поскольку основным рабочим временем пользователя являются рабочие часы с 9 до 18, на графике наблюдается возрастание значений $x(u_e)$ при u_e, принимающих значения из диапазонов [0; 1], [12; 12,5], [19; 24] и убывание значений $x(u_e)$ при u_e, принимающих значения из диапазонов [5; 9], [12,5; 13].

Для более удобной интерпретации данных при формировании обучающей выборки и, при необходимости, её нормирования, представляется целесообразным привести полученные результаты к общей форме записи нечётких подмножеств.

Таким образом, общая форма записи нечетких подмножеств будет иметь следующий вид:

$$A_e = \sum_{u_e=0}^{24} \mu_{Ae}(u)_e / u_e = \sum_{u_e=0}^{0} 0,9 / u_e + \sum_{u_e=1}^{1} 1,0 / u_e + \sum_{u_e=5}^{5} 1,0 / u_e + \sum_{u_e=6}^{6} 0,9 / u_e + \sum_{u_e=8}^{8} 0,5 / u_e + \sum_{u_e=9}^{9} 0 / u_e + \\ + \sum_{u_e=12}^{12} 0 / u_e + \sum_{u_e=13}^{13} 0,1 / u_e + \sum_{u_e=14}^{14} 0 / u_e + \sum_{u_e=18}^{18} 0 / u_e + \sum_{u_e=19}^{19} 0,5 / u_e + \sum_{u_e=23}^{23} 0,7 / u_e + \sum_{u_e=24}^{24} 0,9 / u_e, \quad (6)$$

где: запись вида $\sum_{u_e=0}^{24} \mu_{Ae}(u_e) / u_e$ не предполагает сумму, но предполагает объединение по всем элементам конечного несущего множества значений u_e; в верхней и нижней частях знака суммы $\sum_{u_e=5}^{5} 1,00 / u_e$ указывается значение точки u_e на универсальном множестве U; справа от знака суммы $\sum_{u_e=5}^{5} 1,00 / u_e$ до знака / указывается значение характеристической функции принадлежности μ_{A_e} в точке u_e.

При формировании обучающих выборок представляется целесообразным определить значения, которые могут негативно сказаться на процессе обучения нейронной сети – значения, не позволяющие сделать однозначный вывод о наличии предпосылок к внесению в цепочку недостоверных ПДн. В теории нечётких множеств такие значения определяются точками перехода ($x(u)$ = 0,5). Для функции принадлежности $\mu_{A_e}(u_e)$ такими точками являются u_e = 8 и u_e = 19.

Аналогичным образом предлагается определить числовые значения остальных входных сигналов нейронной сети: набор данных, с которыми работает пользователь; точка доступа пользователя к информационной системе; набор осуществляемых пользователем действий; общая продолжительность проводимых работ.

В состав формируемой нейронной сети было включено три слоя. Количество нейронов в нейронной сети предлагается детерминировать с количеством обучающих пар при помощи следующей формулы:

$$2(m_1 + m_2 + m_3) < L < 10(m_1 + m_2 + m_3), \quad (7)$$

где m_1 – количество нейронов входного слоя, m_2 – количество нейронов скрытого слоя, m_3 – количество нейронов выходного слоя, L – количество обучающих пар. С учётом

применяемого соотношения нейронов в слоях (m_1 во входном, $2m_1+1$ в скрытом и 1 в выходном), формулу (7) можно представить в следующем виде:

$$6m_1 + 4 < L < 30m_1 + 20. \quad (8)$$

При обучении нейронной сети были использованы выборки, полученные из одного распределения, приближенного к реальной действительности. Соотношения аномального поведения к не аномальному в используемом распределении составляло примерно 1 к 20. Для обучения нейронной сети была использована обучающая выборка, состоящая из 300 обучающих пар.

Таким образом, в состав нейронной сети были включены 32 нейрона – 10 во входном слое, 21 в скрытом и 1 в выходном.

В качестве метода обучения предлагается использовать алгоритм обратного распространения ошибки. Данный алгоритм позволяет минимизировать среднеквадратичную ошибку ИНС.

С учетом выбора, сделанного в пользу использования в качестве метода обучения алгоритма обратного распространения ошибки (для гарантированной минимизации среднеквадратичной ошибки) и гиперболических тангенсов в качестве активационных функций (для обеспечения сходимости алгоритма обратного распространения ошибки), выходной сигнал NET будет принимать значения из диапазона от минус единицы до единицы. Таким образом, наличие аномалии в поведении пользователя предлагается интерпретировать выходом NET, равным единице, а отсутствие аномалии выходом NET, равным минус единице.

В качестве обучающей выборки были взяты 300 обучающих пар, выражающих наборы поведенческих характеристик пользователя и формирующих образы, подаваемые на вход нейросети.

В связи с ограниченной областью значений гиперболического тангенса, обучающая выборка была предварительно масштабирована к соответствующему диапазону значений. Нейросеть обучалась на обучающей выборке до достижения заданной среднеквадратичной ошибки.

Валидационная и тестовая выборки состояли из комбинаций по 100 пар.

Для определения качества работы нейросети представляется целесообразным использовать многопараметрическую метрику, включающую:

1) ограничивающие (satisficing) метрики: время исполнения алгоритма – не более 500 мсек; значение гармонически среднего между точностью и полнотой (F1-метрика) – не менее, чем 0,7; значение разброса – не более 1%;

2) оптимизационную (optimizing) метрику – значение смещения не должно превышать 5%.

В процессе обучения возможные выходные характеристики обучающей и валидационной пар состояли из двух классов – аномальное поведение и не аномальное поведение.

Результаты, которые продемонстрировала нейронная сеть с уточнёнными гиперпараметрами после завершения обучения, представлены в таблице 5.

Смещение составило 4%, а разброс (между смещениями валидационной и тестовой выборок) составил 1%.

Таблица 5 – Результаты работы нейронной сети после обучения

Параметр	Значения в выборках		
	Обучающая	**Валидационная**	**Тестовая**
Истинно положительный (True Positive, TP)	14	5	5
Ложноположительный (False Positive, FP)	5	2	3
Ложноотрицательный (False Negative, FN)	1	1	1
Истинно отрицательный (True Negative, TN)	280	92	91
Достоверность (Accuracy = (TP+TN)/(TP+FP+FN+TN))	–	0,98	0,96
Точность (Precission = TP/(TP+FP))	–	~ 0,71	~ 0,63
Полнота (Recall = TP/(TP+FN))	–	~ 0,83	~ 0,83
F1-мера (F1 Score = 2*(Precission*Recall)/(Precission+Recall))	–	~ 0,77	~ 0,77

В пятой главе представлены основные направления апробации проведённых исследований, в частности:

а) основные положения работы докладывались: на Российской научно-технической конференции «Новые информационные технологии в системах связи и управления» (Калуга, 1 июня 2016 г.); на четвёртой конференции по программной инженерии и организации информации «Software Engineering and Information Management (SEIM-2019)» (Санкт-Петербург, 13 апреля 2019 г.);

б) результаты диссертационного исследования: были использованы в учебном процессе ФГАОУ ВО «СПбГУАП»; были реализованы в ПАО «Интелтех» при проведении научно-исследовательской работы «Разработка информационной системы обработки персональных данных в защищённом исполнении» (шифр «Передовик»); были реализованы: в ООО «СИГМА» при выполнении работ по проектированию систем защиты персональных данных, обеспечения анализа и выявлении аномалий в санкционированном поведении пользователей; в ООО «СИГМА» при выполнении работ по проектированию систем защиты информации объектов критической информационной инфраструктуры, в части обеспечения выявления санкционированных, но противоправных команд на ограничение подачи электроэнергии; при проведении эксперимента по анализу поведения пользователей корпоративной локальной вычислительной сети в АО «Кронштадт Технологии»; были использованы при разработке программного обеспечения, предназначенного для проведения анализа санкционированного поведения пользователей; были использованы при подготовке патентной документации на полезную модель «Автоматизированная система управления защитой информации в территориально распределённой информационной системе».

В заключении сделаны выводы, что поставленная в работе цель (обеспечение достоверности данных при их обработке в блокчейн-системе) достигнута, а задачи

решены. Достижение поставленной цели подтверждается появлением в системе защиты данных, обрабатываемых в блокчейн-системе, новых функций и свойств:

1) при выявлении актуальных угроз нарушения информационной безопасности данных, обрабатываемых в блокчейн-системе, учитываются: специфические угрозы данным, обрабатываемым в блокчейн-системе; числовые значения степени опасности нарушения отдельных характеристик безопасности (в т.ч. достоверности); зависимость между: угрозами, актуальными для данных, обрабатываемых в блокчейн-системе; ущербом от потенциальных угроз блокчейн-системе; степенью опасности нарушения отдельных характеристик безопасности (в т.ч. достоверности); составом деструктивных воздействий; степенью важности данных;

2) при обработке данных в блокчейн-системе осуществляется: автоматизированная оценка рисков внесения и обработки недостоверных данных; анализ санкционированного поведения пользователей и выявление аномалий в их поведении.

При этом затраты на внедрение и сопровождения предложенного автором решения по анализу санкционированного поведения пользователей и выявлению аномалий ниже, чем у аналога. Требуемый результат при этом достигается в полной мере.

Основные результаты диссертационной работы сводятся к следующему:

1. Проведён анализ существующих требований и методов разработки информационных систем и защиты данных на примере Российских нормативных актов об ИСПДн и ПДн, как соответствующих общим международным требованиям.

На основании проведённого анализа существующих требований и методов обеспечения безопасности данных обоснована актуальность разработки метода обеспечения безопасности данных при их обработке в блокчейн-системе за счёт применения искусственных нейронных сетей.

2. Разработана модель выявления актуальных угроз нарушения информационной безопасности данных. В рамках разработки модели установлена и формализована зависимость между: угрозами, актуальными для данных, обрабатываемых в блокчейн-системе; ущербом от потенциальных угроз блокчейн-системе; степенью опасности нарушения отдельных характеристик безопасности (в т.ч. достоверности); составом деструктивных воздействий; степенью важности данных, обрабатываемых в блокчейн-системе.

Предложенная модель: 1) отличается от классической представлением угроз, учитывающим: специфические угрозы данным, обрабатываемым в блокчейн-системе; числовые значений степени опасности нарушения отдельных характеристик безопасности (в т.ч. достоверности); 2) позволяет: определить угрозы, актуальные для блокчейн-систем; определить угрозы, оказывающие влияние на достоверность данных; в частности, в России – обеспечить выполнение требований государственных стандартов в части определения актуальных угроз.

3. Разработан метод обеспечения достоверности ПДн, обрабатываемых в блокчейн-системе. В рамках разработки метода расширен класс методов обеспечения

достоверности данных в части выявления недостоверных ПДн при их вводе в блокчейн-систему за счёт искусственной нейронной сети.

Предложенный метод: 1) отличается от известных концептуально новым подходом к достижению консенсуса, включающим процедуру автоматизированной оценки рисков внесения и обработки недостоверных данных, разработанную с использованием теории искусственных нейронных сетей и теории нечетких множеств; 2) позволяет обеспечить достоверность ПДн при их обработке в блокчейн-системе: основанных на облачных платформах, как на уровне организаций различных форм собственности, так и на уровне государства в целом; в частности в России – в рамках обеспечения выполнение требований ФОИВ ТЗИ в части контроля ошибочных действий пользователей по вводу данных.

4. Разработана методика анализа санкционированного поведения пользователей информационной системы. В рамках разработки методики формализовано поведение пользователя и доказана возможность выявления аномалий в поведении пользователя при помощи искусственных нейронных сетей.

Предложенная методика: 1) отличается от известных уникальным составом формализованных характеристик поведения пользователя (составом входных характеристик нейронной сети) и параметрами нейронной сети; 2) позволяет оперативно выявлять аномалии в поведении пользователей, а также при применении в России: обеспечить выполнение требований ФОИВ ТЗИ в части анализа зарегистрированных событий безопасности и реагирования на них; при защите от угроз, представленных в банке данных угроз ФОИВ ТЗИ, связанных с подменой доверенного пользователя и его действий путём обмана.

Эксперимент показал, что смещение нейронной сети, разработанной в рамках предложенной методики, при эксплуатации в условиях, максимально приближенных к реальным, составило порядка 4,3 %, что является достаточно хорошим результатом.

5. Реализация результатов проведённого исследования подтверждена актами внедрения ООО «СИГМА и ФГАОУ ВО «СПб ГУАП».

6. Дальнейшие научные исследования по теме диссертации представляется целесообразным продолжить в следующих направлениях: расширение состава специфических угроз, актуальных для блокчейн-систем; разработка более точных методов присвоения числовых значений факторам, создающим предпосылки для внесения и обработки в блокчейн-системе недостоверных данных; разработка более точных методов присвоения числовых значений характеристикам санкционированного поведения пользователей; исследование альтернативных конфигураций нейронных сетей, предназначенных для обеспечения анализа рисков внесения и обработки в блокчейн-системе недостоверных данных; исследование альтернативных конфигураций нейронных сетей, предназначенных для обеспечения выявления аномалий в санкционированном поведении пользователей.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

Публикации в рецензируемых научных изданиях, рекомендованных ВАК:

1) Козин И.С. Метод обеспечения безопасности персональных данных при их обработке в информационной системе при помощи искусственной нейронной сети // Информатизация и связь. 2021. № 4. С. 51–56.

2) Козин И.С. Метод обеспечения безопасной обработки персональных данных на основе применения технологии блокчейн // Научно-технический вестник информационных технологий, механики и оптики. 2019. Т. 19. № 5. С. 892–900 doi: ISSN: 2226-1494.

3) Козин И.С. Метод обеспечения безопасности персональных данных при их обработке в информационной системе на основе анализа поведения пользователей// Информационно-управляющие системы. 2018. № 3. С. 69–78. doi:10.15217/issn1684-8853.2018.3.69.

4) Козин И.С. Метод разработки автоматизированной системы управления информационной безопасностью распределённой информационной системы // Информация и космос. 2018. № 3. С. 80–88.

5) Козин И.С., Беззатеев С.В. Метод определения опасности угрозы персональным данным при их обработке в информационной системе // Известия СПбГЭТУ «ЛЭТИ». 2017. № 10. С. 19–26.

Опубликованные в других изданиях:

6) Козин И.С., Рощин А.А. Метод обеспечения безопасности информации при её обработке в информационной системе на основе машинного обучения // Техника средств связи. 2019. № 4 (148). С. 70–82.

7) Kozin I.S. Providing personal data protection based on the block chain technology // Fourth Conference on Software Engineering and Information Management (SEIM-2019) (Saint-Petersburg, April 13, 2019). P. 10–16.

8) Козин И.С., Рощин А.А. Метод определения опасности угрозы персональным данным личного состава объекта // Техника средств связи: науч.-техн. сб. СПб.: Изд-во Политех. ун-та, 2017. № 6. С. 123–131.

9) Козин И.С. Метод разработки автоматизированной системы управления информационной безопасностью региональной информационной системы // Сборник трудов «Региональная информатика и информационная безопасность. Выпуск 3» Санкт-Петербург. СПОИСУ. СПб., 2017. С. 283–289.

10) Козин И.С., Рощин А.А. Метод построения модели угроз критически важной информации военного назначения // Техника средств связи: науч.-техн. сб. СПб.: Изд-во Политех. у-та, 2016. №5. С. 98–103.

Отчёт о НИР:

11) Козин И.С., Гузарев А.С., Баринов Н.О., Бугаенко И.И., Ветров И.М., Иванцова И.Е., Малаховский А.А. Отчёт о научно-исследовательской работе «Разработка информационной системы обработки персональных данных в защищённом исполнении» (заключительный). 2017. 135 с.

Подписано в печать 31.01.2022. Формат 60×84 1/16.
1,0 а.л. Тираж 100 экз.

Отпечатано в СПбГУТ, 193232, Санкт-Петербург, пр. Большевиков, д. 22, корп. 1