УНИВЕРСИТЕТ ИТМО
НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
сентябрь–октябрь 2019 Том 19 № 5 ISSN 2226-1494 http://ntv.itmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTCS
September–October 2019 Vol. 19 No 5 ISSN 2226-1494 http://ntv.itmo.ru/en/
НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК УНИВЕРСИТЕТ ИТМО ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

УДК 004.056 doi: 10.17586/2226-1494-2019-19-5-892-900

МЕТОД ОБЕСПЕЧЕНИЯ БЕЗОПАСНОЙ ОБРАБОТКИ ПЕРСОНАЛЬНЫХ ДАННЫХ НА ОСНОВЕ ПРИМЕНЕНИЯ ТЕХНОЛОГИИ БЛОКЧЕЙН

И.С. Козин

АО «Кронштадт Технологии», Санкт-Петербург, 199178, Российская Федерация
Адрес для переписки: ivan.kozin@krontech.ru

Информация о статье
Поступила в редакцию 25.05.19, принята к печати 17.07.19
Язык статьи — русский

Ссылка для цитирования: 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: 10.17586/2226-1494-2019-19-5-892-900

Аннотация
Разработан подход к созданию распределенной информационной системы обработки персональных данных, основанной на технологии блокчейн, включающий предложения по определению общей архитектуры системы, а также порядка хранения данных, вознаграждения пользователей, достижения консенсуса, внедрения и развития системы. Хранение данных обеспечивается с применением личных устройств пользователей, а также средств криптографической защиты информации. Механизм вознаграждения пользователей основан на применяемой в Китае системе социального кредитования, обеспечивающей отбор наиболее благонадежных субъектов персональных данных, способных занять роль узлов консенсуса. Процедура достижения консенсуса включает в себя автоматизированный анализ рисков обработки недостоверных данных. В качестве математических аппаратов анализа рисков предложены теория искусственных нейронных сетей и теория нечетких множеств. Применение искусственной нейронной сети обеспечивает гибкость системы в целом в условиях роста количества пользователей. Применение предложенного подхода к созданию распределенной информационной системы позволит обеспечить повышение доступности, целостности и конфиденциальности данных за счет децентрализованной обработки, а также применения хорошо изученных методов криптографической защиты.

Ключевые слова
информационная безопасность, персональные данные, блокчейн, децентрализация, достижение консенсуса, социальное кредитование

doi: 10.17586/2226-1494-2019-19-5-892-900

BLOCKCHAIN-BASED TECHNOLOGY FOR SECURITY PROCESSING OF PERSONAL DATA

I.S. Kozin

JSC Kronshtadt Technologies, Saint Petersburg, 199178, Russian Federation
Corresponding author: ivan.kozin@krontech.ru

Article info
Received 25.05.19, accepted 17.07.19
Article in Russian

For citation: Kozin I.S. Blockchain-based technology for security processing of personal data. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2019, vol. 19, no. 5, pp. 892–900 (in Russian). doi: 10.17586/2226-1494-2019-19-5-892-900

Abstract
The paper presents the developed approach of creating a decentralization personal data information system, based on the blockchain technology. The approach includes proposals for overall system architecture specification, data storage procedure, users' fee, consensus mechanism, and system implementation and enhancement. Data storage is provided with the use of personal users' devices and information cryptographic protection facilities. The users' fee mechanism is based on the social credit system, employed in China, which ensures the selection of the most trustworthy personal data subjects able to play the role of consensus nodes. Consensus procedure includes automated risk analysis of unreliable data processing. A neural network theory and fuzzy set theory are proposed as the mathematical tools of risk analysis. The use of an artificial neural network provides flexibility of the system as a whole in terms of the growing number of users. The proposed approach application for decentralization information system design will provide for improvement of availability, integrity and confidentiality of data through decentralized processing and application of well-studied cryptographic protection methods.

Keywords
information security, personal data, blockchain, decentralization, consensus mechanism, social credit

Введение

В условиях активного развития информационных технологий все более значительное место в развитии общества занимает информационная безопасность. Обеспечение безопасности информации при ее обработке в информационных системах является высокорентабельным бизнесом, отличающимся высоким рыночным потенциалом и эффективностью инвестиций.

С учетом положений Указа Президента[1] в качестве одного из основных направлений обеспечения информационной безопасности можно выделить защиту персональных данных (ПДн). В соответствии с Федеральным законом[2] примерами таких данных являются паспортные данные, сведения о доходах, состоянии здоровья, политических взглядах, религиозных и философских убеждениях. Нарушение безопасности ПДн может привести к материальному и моральному ущербу как для субъекта ПДн, так и для организации, осуществляющей обработку ПДн. Таким образом, обеспечение безопасности ПДн при их обработке в информационной системе является важной и актуальной задачей.

В настоящее время существует множество подходов к обеспечению безопасности информации при ее обработке в информационной системе. К недостаткам классических подходов относится недостаточная проработка решений по обеспечению доступности и целостности данных при репликации: значительное повышение вероятности зависаний системы с увеличением числа узлов [1]; возможность отказа и потери данных [2]; высокие показатели сложности разрешения конфликтов, времени синхронизации, а также сложности реализации [3]; значительное увеличение нагрузки на магистральные каналы передачи данных при выходе из строя аппаратного обеспечения централизованных хранилищ (кластеров), высокая сложность планирования проведения регламентных работ по обслуживанию инфраструктуры [4].

Одним из подходов к повышению безопасности информации является распределенная обработка, основанная на применении технологии цепочки блоков — блокчейн (англ. block chain).

Блокчейн — это технология обработки данных, в основе которой можно выделить следующие основные принципы: структура хранения данных представляет собой выстроенную по определенным правилам цепочку блоков, содержащих информацию; каждый блок цепочки связан с соседними блоками криптографическими методами; сами по себе блоки и содержащаяся в них информация являются общедоступными; копии цепочки блоков хранятся на разных компьютерах.

Примерами решений, построенных на использовании технологии блокчейн, и предназначенных для обработки ПДн, являются проекты DACC[3], Goldilock[4], Decenturion[5].

Постановка задачи построения распределенного реестра обработки персональных данных

В целом отличительными характеристиками современных информационных систем персональных данных (ИСПДн) являются: наличие одного или нескольких централизованных хранилищ данных, вывод из строя которых может привести к значительному нарушению доступности ПДн; ограниченные возможности оперативного доступа к ПДн.

В настоящей статье предлагается метод обеспечения безопасности ПДн при их обработке (в том числе накоплении, хранении, передаче и т.п. в соответствии с Федеральным законом «О персональных данных») в распределенном реестре, основанном на технологии блокчейн, и обеспечивающий повышенные доступность и целостность ПДн за счет отсутствия подверженного отказу единого центра обработки данных; повышенную доступность за счет распределенной обработки ПДн и объединения разнородных ПДн в единую взаимоувязанную систему. Несмотря на активное развитие технологии «блокчейн», в том числе в области обработки и защиты персональных данных, методы и подходы к ее применению на сегодняшний день изучены не в полной мере.

Построение распределенного реестра обработки ПДн (РРПДн) сводится к решению следующих задач: определение состава ПДн, обработку которых целесообразно осуществлять в распределенной системе; определение общей архитектуры РРПДн; определение порядка хранения данных; определение порядка вознаграждения пользователей РРПДн, обеспечивающих его работу; определение механизма достижения консенсуса; выбор способа вычисления хеш-функции.

---

[1] Указ Президента РФ от 06.03.1997 № 188 «Об утверждении перечня сведений конфиденциального характера». http://www.kremlin.ru/acts/bank/10638 (дата обращения: 31.01.2019).

[2] Федеральный закон от 27.07.2006 № 152-ФЗ «О персональных данных». https://rg.ru/2006/07/29/personaljnye-dannye-dok.html (дата обращения: 31.01.2019).

[3] URL: https://dacc.com/ (дата обращения: 22.03.2019)

[4] URL: https://www.goldilock.com/ (дата обращения: 22.03.2019)

[5] URL: https://decenturion.com/ (дата обращения: 22.03.2019)

Для примера в качестве района эксплуатации РРПДн предлагается рассмотреть условный субъект Российской Федерации (РФ), численность населения которого составляет 1 млн человек.

Построение распределенного реестра обработки персональных данных

Состав обрабатываемых персональных данных. В качестве примера предлагается рассмотреть РРПДн, предназначенный для обеспечения оперативного доступа к ПДн и решения следующих задач: обеспечение идентификации личности; обеспечение доступа к сведениям об образовании и профессиональных навыках; создание и обмен умными активами; создание и выполнение умных контрактов.

Для решения указанных задач в РРПДн необходимо обрабатывать следующие сведения: содержащиеся в паспорте гражданина РФ; об образовании и профессиональных навыках; о финансовом состоянии. На сегодняшний день основными операторами указанных ПДн являются органы государственной власти (МВД России, ФМС России и т.п.), учебные заведения и банковские учреждения.

Общая архитектура распределенного реестра. Предлагается выделить в РРПДн несколько самостоятельных цепочек блоков (ЦБ) по одной для каждой предметной области (ЦБ-И для идентификационных данных, ЦБ-П для данных о профессиональных навыках, ЦБ-А для данных об активах и ЦБ-К для данных об умных контрактах).

Узлы РРПДн предлагается разделить на три типа: узлы консенсуса, узлы аудита и легкие клиенты. Узлы консенсуса должны принимать участие в формировании новых блоков, внося ПДн в блоки и распространяя их по сети. Узлы аудита должны содержать копию цепочки блоков (или ее фрагментов) и обеспечивать распределение нагрузки по сети, в совокупности выполняя роль сети доставки контента (Content Delivery Network, CDN), т.е. обеспечивать: передачу данных между легкими клиентами и узлами консенсуса; снижение количества транзитных участков; предотвращение задержек, прерываний связи и потери на перегруженных каналах и стыках между ними. Легкие клиенты предназначены для установки на платформах с невысокими показателями производительности (в том числе мобильных) и могут содержать только данные, необходимые для конкретного узла.

Частота обновления цепи (создания новых блоков) должна быть обусловлена спецификой области обработки ПДн. Данные ЦБ-И предлагается обновлять один раз в день, данные ЦБ-П — один раз в неделю, данные о внесении значительных изменений в сведения о финансовом состоянии, а также выполнении контрактов — один раз в час.

Ядро платформы РРПДн предлагается реализовать на языке программирования Java 8 с применением NoSQL базы данных. Взаимодействие с ядром платформы предлагается построить на базе архитектуры RESTful API. Подобный подход к построению архитектуры распределенного реестра, предназначенного для обработки ПДн, реализован в платформе Erachain[1]. При разработке РРПДн рекомендуется использовать методы разработки безопасного программного обеспечения (например, предложенные в работах [5, 6]). Также при проектировании РРПДн может найти применение методика расчета надежности сложных систем, предложенная в работе [7].

Порядок хранения данных. Объемные данные предлагается хранить на личных носителях информации пользователей РРПДн (субъектов ПДн), являющихся узлами аудита или консенсуса. При необходимости обработки биометрических данных может найти применение метод маскирующего сжатия на основе модели взвешенной структуры изображения, представленный в работе [8]. Для обеспечения конфиденциальности ПДн предлагается обеспечить их шифрование.

Распределенное хранилище ПДн должно обеспечивать хранение больших по объему файлов (таких как документы Microsoft Word, биометрические данные). Примером системы, которая способна обеспечить работу РРПДн, является проект InterPlanetary File System (IPFS).

Также предлагается заложить в блоки цепи условие архивирования неиспользуемых блоков. Для архивирования могут применяться такие системы, как Internet Archive, Wayback Machine или аналогичные.

С учетом численности населения района эксплуатации (1 млн пользователей — субъектов ПДн) примерный объем актуальных данных РРПДн составит до 29 ГБ (из расчета: 1,5 КБ на одного субъекта в ЦБ-И, 25 КБ на одного субъекта ЦБ-П, по 1 КБ на одного субъекта ЦБ-А и ЦБ-К).

Порядок вознаграждения. Для стимулирования субъектов ПДн к участию в обеспечении работы РРПДн предлагается ввести в РРПДн три учетные единицы вознаграждения — баллы социальной надежности (БСН), баллы социальной эффективности (БСЭ) и социальные монеты (СМ).

Социальные баллы (БСН и БСЭ), по определению автора, это неотчуждаемая характеристика субъекта, выражающая общую пользу, которую субъект приносит обществу. БСН предлагается начислять за достоинства, формирующие фундамент доверия к пользователю РРПДн (отсутствие правонарушений, вовремя выплачиваемые кредиты и прочие, в том числе за регистрацию в РРПДн и хранение данных РРПДн на личном устройстве), а БСЭ — за деятельность, способствующую развитию общества в целом (выдающиеся успехи в профессиональной деятельности, продвижение инновационных технологий и т. п.). Подобный

[1] URL: https://www.erachain.org/home (дата обращения: 22.03.2019)

подход к построению общественных отношений в настоящее время (в период с 2014 по 2020 г.) активно внедряется в Китае в рамках программы «Система социального кредитования» [9]. Членство в списке «заслуживающих доверие» дает повышенный приоритет при оказании различных услуг. Попадание в список «недобросовестных» влечет за собой различные ограничения.

Вознаграждение за создание новых блоков можно реализовать средствами социальных монет, в перспективе подлежащих обмену на фиатные деньги. При этом возможность участия в создании новых блоков (с получением вознаграждения), а также вероятность успеха должны зависеть от количества БСН и БСЭ пользователя (узла консенсуса).

Подход к вознаграждению, а также определение его размера требует серьезной проработки. В табл. 1 представлен обобщенный пример, показывающий возможные значения вознаграждений. Для обоснования числовых значений вознаграждений может быть применена теория нечетких множеств. В рамках предложенного примера узлами консенсуса могут становиться пользователи, набравшие суммарно 8 социальных баллов.

Таблица 1. Обобщенный пример возможных значений вознаграждений

Цепочка	Основание для вознаграждения	Размер вознаграждения		
		БСН	БСЭ	СМ
Все	Регистрация	1		
Все	Хранение данных (от 10 до 100 % цепочки)	0,1–0,5		
Все	Создание новых блоков			1
ЦБ-П	Получение образования (среднее/среднее-специальное/высшее)	0,5/1/2		
ЦБ-П	Повышение квалификации/профессиональная переподготовка	0,5/1		
ЦБ-П	Получение ученой степени (кандидат наук/доктор наук)	3/6		
ЦБ-П	Активное участие в создании до 10/100/1000 рабочих мест		1/3/6	
ЦБ-А	Накопление активов на сумму в 10/100/1000 млн рублей	3/6/9		
ЦБ-А	Успешное погашение кредитных обязательств на сумму до 100/1000/10000 тыс. рублей	0,1/0,5/2		
ЦБ-А	Невыполнение кредитных обязательств на сумму до 100/1000/10000 тыс. рублей	–1/–2/–5		
ЦБ-К	Успешное выполнение контракта на сумму до 10/100/1000 млн рублей	1	5/10/20	
ЦБ-К	Невыполнение контрактных обязательств на сумму до 10/100/1000 млн рублей	–1	–5/–10/–20	

Достижение консенсуса. Процедура достижения консенсуса заключается в подтверждении вносимых в реестр ПДн и состоит из следующих шагов:

— передача пользователем своих ПДн (в том числе копий официальных документов для подтверждения достоверности) в очередь транзакций с использованием хорошо изученных методов шифрования;

— подтверждение достоверности ПДн, находящихся в очереди транзакций (вручную на основе копий официальных документов, а также с помощью автоматизированной оценки рисков);

— формирование нового блока.

Существует вероятность внесения и обработки в РРПДн недостоверных ПДн. Мотивом для внесения в реестр таких данных могут служить стремление субъекта к финансовой выгоде, высокому социальному статусу и т. п. Для повышения достоверности вносимых в реестр данных предлагается реализовать механизм автоматизированной оценки рисков внесения и обработки недостоверных ПДн, основанный на использовании методов машинного обучения. В качестве исходных (входных) данных предлагается выделить факторы, создающие предпосылки для внесения и обработки в РРПДн недостоверных ПДн. Такими факторами могут быть: повышенная вероятность вступления в сговор субъекта и объекта подтверждения; возможность получения значительного вознаграждения; невысокая степень надежности объекта подтверждения (выраженная небольшим количеством социальных баллов и отражающая в том числе невысокий уровень материального благосостояния, отсутствие востребованных знаний и навыков и т. п.).

В качестве примера предлагается выделить четыре характеристики, с использованием которых может быть проведен анализ рисков:

1) степень формальной связи узла консенсуса (пользователя, осуществившего ручное подтверждение достоверности ПДн, передаваемых в РРПДн) и объекта подтверждения, которая может быть обусловлена общим учебным заведением или местом работы, родственной связью, участием в общих умных контрактах и т. п.;

2) степень участия в сети взаимных подтверждений ПДн (может свидетельствовать об осуществлении сговора среди пользователей и совершении действий, распределенных среди нескольких человек и от того особо затруднительных в выявлении);

3) размер потенциального вознаграждения субъекта ПДн;

4) степени надежности узла консенсуса и объекта подтверждения.

Таким образом, в рассматриваемом примере риск обработки недостоверных ПДн будет выражен риском вступления в сговор узла консенсуса и объекта подтверждения. Каждую из представленных характеристик можно описать в виде коэффициентов x_n, $n \in \{1; 4\}$, где n выражает порядковый номер характеристики: x_1 — степень связи узла консенсуса и объекта подтверждения; x_2 — степень участия в сети взаимных подтверждений ПДн; x_3 — размер потенциального вознаграждения субъекта ПДн; x_4 — степень надежности узла консенсуса и объекта подтверждения.

В качестве математического аппарата оценки риска при подтверждении ПДн предлагается использовать теорию искусственных нейронных сетей (ИНС). Таким образом, на входе нейронной сети должно быть четыре входных сигнала x_1–x_4, и построение ИНС сводится к решению следующих задач: определение типа необходимой ИНС; определение подхода к присвоению числовых значений входным сигналам ИНС, выражающим анализируемые характеристики (x_1–x_4); определение необходимого количества слоев ИНС и количества нейронов в слоях ИНС; выбор метода обучения ИНС; выбор активационных функций; выбор области значений выходного сигнала NET, сигнализирующего о степени риска при подтверждении ПДн. Значимость входных значений будет определена в ходе обучения ИНС путем изменения весовых коэффициентов нейронных связей.

Поскольку на сегодняшний день не существует строгой теории по выбору ИНС [10], 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. Подобный подход описан в работе [11].

Из проведенных А.Н. Колмогоровым исследований [12] можно сделать вывод о том, что любую непрерывную функцию f:$[0;1]^n \rightarrow [0;1]$ можно аппроксимировать при помощи трехслойной нейронной сети, имеющей n входных, $2n + 1$ скрытых и один выходной нейрон. Таким образом, при построении ИНС предлагается использовать три слоя: первый слой будет включать в себя два нейрона, второй слой — пять нейронов, последний слой (выходной) — один нейрон.

В качестве метода обучения предлагается использовать алгоритм обратного распространения ошибки [13]. Данный алгоритм позволяет минимизировать среднеквадратичную ошибку ИНС.

Для минимизации среднеквадратичной ошибки ИНС изменение весовых коэффициентов и порогов нейронов необходимо осуществлять в соответствии со следующими выражениями:

$$w_{ij}(t + 1) = w_{ij}(t) - \alpha\gamma_j F'(S_j)y_i,$$
$$T_j(t + 1) = T_j(t) - \alpha\gamma_j F'(S_j),$$

где w_{ij} — это значение веса j-го нейрона слоя i.

Для обеспечения сходимости алгоритма обратного распространения ошибки в качестве активационной функции предполагается использовать гиперболический тангенс. Таким образом, ИНС будет являться гомогенной, а выходное значение j-го нейрона определяться следующим образом:

$$y = \mathrm{tg}(S_j) = \frac{e^{S_j} - e^{-S_j}}{e^{S_j} + e^{-S_j}},$$

где S_j — взвешенная сумма j-го нейрона. Поскольку производная этой функции имеет вид $F'(S_j) = 1 - y_j^2$, правило обучения можно представить в виде:

$$w_{ij}(t + 1) = w_{ij}(t) - \alpha\gamma_j(1 - y_j^2)y_i,$$
$$T_j(t + 1) = T_j(t) + \alpha\gamma_j(1 - y_j^2),$$

где t — номер итерации, α — значение шага обучения, γ_i — значение ошибки для i-го нейрона, T_j — значение порога j-го нейрона.

Ошибка для j-го нейрона выходного и скрытого слоев определяется следующим образом:

$$\gamma_j = y_j - t_j,$$
$$\gamma_j = \sum_i \gamma_i(1 - y_i^2)w_{ij}.$$

Описание порядка выполнения алгоритма обратного распространения ошибки, а также общие рекомендации представлены в работе [14].

С учетом выбора, сделанного в пользу использования в качестве метода обучения алгоритма обратного распространения ошибки и, как следствие, гиперболических тангенсов в качестве активационных функций, выходной сигнал NET будет принимать значения из диапазона от минус единицы до еди-

ницы. Таким образом, наличие сговора между объектом подтверждения и узлом консенсуса предлагается интерпретировать выходом NET, равным единице, а отсутствие сговора выходом NET, равным минус единице.

В качестве итоговой конфигурации был выбран трехслойный полносвязный гомогенный персептрон без обратных связей с четырьмя входными сигналами, восемью нейронами и гиперболическими тангенсами в качестве функций активации. В первом слое содержится два нейрона, в скрытом слое — пять и в выходном — один. Итоговая конфигурация подготовленной ИНС представлена на рисунке.

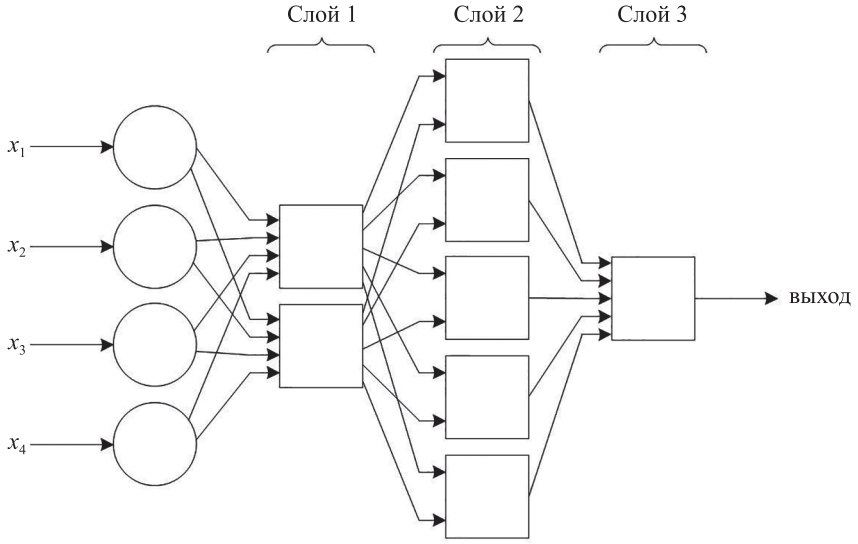

Рисунок. Итоговая конфигурация искусственной нейронной сети

Обучение полученной ИНС проводилось в программном обеспечении SPSS Statistics (разработка компании IBM), предназначенном для статистической обработки данных. После обучения нейронной сети была проведена проверка эффективности ее работы с помощью контрольной выборки. Относительная погрешность классификации данных составила примерно 10 %, что является достаточно хорошим результатом. При обучении использовались небольшие выборки, предназначенные для демонстрации общих принципов работы ИНС. Для применения представленного подхода в решении реальных прикладных задач выборки должны быть больше. После ввода РРПДн в эксплуатацию должно осуществляться регулярное обучение подсистемы оценки рисков (обучение ИНС без учителя). Аналогичный подход к построению ИНС ранее предлагался в работе [11], посвященной выявлению аномалий в санкционированном поведении пользователей.

Обобщенный порядок внедрения и развития. Поскольку подтверждение данных, передаваемых в распределенный реестр, их хранение, а также формирование новых блоков предполагается осуществлять силами пользователей РРПДн, предлагается выделить две стадии развития реестра. На первой стадии необходимо заложить надежную основу реестра, состоящую из наиболее благонадежных пользователей (например, работников научно-исследовательских учреждений). Такой подход обеспечит повышение достоверности данных, которые будут вноситься в реестр при его дальнейшем развитии. На второй стадии предлагается расширить круг лиц, которым предоставляется право регистрации в РРПДн, до всех жителей района эксплуатации.

Первая стадия развития распределенного реестра

В рамках первой стадии предлагается выделить три этапа:

1) формирование ЦБ-И как основы РРПДн, содержащей уникальные идентификаторы пользователей;

2) формирование ЦБ-П как основы для создания конкуренции в вычислительной среде пользователей РРПДн;

3) формирование ЦБ-А и ЦБ-К.

Поскольку идентификационные данные субъекта являются ключевыми данными РРПДн, регистрация в ЦБ-И должна осуществляться на основании документа, удостоверяющего личность гражданина РФ. При этом вновь зарегистрированные пользователи получают уникальный идентификатор (УИД) и равное количество БСН за регистрацию. Формирование новых блоков на этом этапе должно осуществляться силами разработчика РРПДн. Сразу после регистрации пользователю должна предоставляться возможность выбора степени своего участия в работе ЦБ-И:

— простая регистрация;

— предоставление услуг хранения распределенной базы данных ЦБ-И.

В первом случае пользователь просто переносит свои ПДн в распределенный реестр и получает доступ к функционалу РРПДн посредствам легкого клиента. Во втором случае пользователь, выбирая приемлемый объем данных, который он готов хранить на личном устройстве, становится узлом аудита. Размер вознаграждения за хранение данных (дополнительные БСН) зависит от выбранного пользователем объема данных.

Регистрация пользователей и передача ПДн в ЦБ-П должны осуществляться с использованием УИД ЦБ-И. Переданные в ЦБ-П ПДн пользователя (сведения об образовании, профессиональных навыках) должны быть подтверждены соответствующими официальными документами. Вновь зарегистрированные пользователи получают равное количество БСН за регистрацию, а также дополнительные БСН и БСЭ, количество которых зависит от достижений пользователя в профессиональной деятельности (в том числе образовательной, научно-исследовательской и т. п.). Формирование новых блоков должно осуществляться силами разработчика РРПДн до достижения определенного количества пользователей, обладающих необходимым количеством БСН и БСЭ. После этого таким пользователям должна быть предоставлена возможность стать узлами консенсуса и принять участие в создании новых блоков. Возможность формирования нового блока (с получением вознаграждения в виде социальных монет) должна зависеть от количества БСН и БСЭ, которыми обладает узел консенсуса. Процедура ознакомления узлов консенсуса с вносимыми в реестр ПДн пользователей может быть проведена двумя способами, выбор между которыми должен предоставляться субъекту ПДн:

— субъект сам определяет критерии узлов консенсуса, которые могут получить доступ к его ПДн с целью подтверждения (работники конкретной организации, владельцы конкретных УИД и т. п.);

— пользователь делает свои ПДн общедоступными для узлов консенсуса на время подтверждения.

Формирование ЦБ-А и ЦБ-К предлагается начать после достижения необходимого количества узлов консенсуса в среде пользователей РРПДн. Регистрация и передача ПДн в ЦБ-А и ЦБ-К должны осуществляться с использованием УИД ЦБ-И. Переданные в ЦБ-А сведения о находящихся в собственности активах должны быть подтверждены соответствующими официальными документами. Подтверждение данных, передаваемых в ЦБ-К, предлагается осуществлять без предъявления официальных документов. При этом пользователю, по аналогии с процедурой подтверждения ЦБ-П, должна предоставляться возможность «фильтрации» узлов консенсуса, которые могут подтвердить передаваемые данные. Вновь зарегистрированные пользователи получают равное количество БСН за регистрацию.

В качестве альтернативного подхода или в качестве меры, повышающей достоверность вносимых ПДн, подтверждение данных может осуществляться с участием уполномоченных органов государственной власти.

Вторая стадия развития распределенного реестра

Расширение круга лиц, которым предоставляется право регистрации в РРПДн на второй стадии, предлагается осуществлять постепенно, предоставляя права регистрации в первую очередь группам лиц, в силу объективных причин заслуживающих наибольшее доверие (например, работникам органов государственной власти, студентам наиболее заслуженных учебных заведений и т. п.). В результате право регистрации в РРПДн должно быть предоставлено всем жителям района эксплуатации (гражданам РФ). В целом развитие РРПДн на второй стадии предлагается реализовать в порядке, аналогичном порядку первой стадии: сначала пользователи регистрируются в ЦБ-И (получают УИД), затем получают право регистрации в других цепочках. Обобщенные сведения о предложенном порядке развития РРПДн представлены в табл. 2.

Таблица 2. Обобщенный порядок развития РРПДн

Состав субъектов	Механизм подтверждения ПДн	Порядок создания блоков	Порядок вознаграждения	Результат
Стадия 1. Формирование базы данных наиболее благонадежных пользователей				
Этап 1. Формирование ЦБ-И				
Ограниченный круг доверенных лиц	Предъявление документа, удостоверяющего личность	Силами разработчика РРПДн	БСН за регистрацию и хранение	Формирование ЦБ-И с наиболее доверенными пользователями
Этап 2. Формирование ЦБ-П				
Ограниченный круг доверенных лиц	1) Предъявление официальных документов; 2) Подтверждение пятью узлами консенсуса достоверности ПДн и блоков; 3) Автоматизированная оценка рисков обработки недостоверных ПДн	1) Силами разработчика РРПДн; 2) Силами узлов консенсуса	1) БСН за регистрацию и хранение; 2) БСН и БСЭ за достижения; 3) Социальные монеты за создание новых блоков	Формирование ЦБ-П и необходимого количества узлов консенсуса

Таблица 2. Продолжение

Состав субъектов	Механизм подтверждения ПДн	Порядок создания блоков	Порядок вознаграждения	Результат
Этап 3. Формирование ЦБ-А и ЦБ-К				
Ограниченный круг доверенных лиц	1) Предъявление официальных документов (только для ЦБ-А); 2) Подтверждение пятью узлами консенсуса достоверности ПДн и блоков; 3) Автоматизированная оценка рисков обработки недостоверных ПДн	Силами узлов консенсуса	1) БСН за регистрацию и хранение; 2) БСН и БСЭ (в ЦБ-П и ЦБ-К) за достижения; 3) Социальные монеты за создание новых блоков	Формирование ЦБ-А и ЦБ-К
Стадия 2. Расширение круга лиц, которым предоставляется право регистрации в РРПДн				
Все жители района эксплуатации	1) Предъявление официальных документов (только для ЦБ-И, ЦБ-П и ЦБ-А); 2) Подтверждение пятью узлами консенсуса достоверности ПДн и блоков; 3) Автоматизированная оценка рисков обработки недостоверных ПДн	Силами узлов консенсуса	1) БСН за регистрацию и хранение; 2) БСН и БСЭ (в ЦБ-П и ЦБ-К) за достижения; 3) Социальные монеты за создание новых блоков	Перенесение в распределенный реестр ПДн всех жителей района эксплуатации (граждан РФ)

Заключение

Предложен метод обеспечения безопасности персональных данных при их обработке в распределенном реестре, основанный на использовании технологии блокчейн. Предложенный метод включает предложения по построению общей архитектуры распределенного реестра, порядку хранения данных, вознаграждению участников работы распределенного реестра, способу достижения консенсуса, а также обобщенному порядку внедрения и развития системы.

Метод отличается от известных уникальной архитектурой информационной системы персональных данных; способом достижения консенсуса, включающим процедуру автоматизированной оценки рисков внесения и обработки недостоверных персональных данных, основанную на применении теории искусственных нейронных сетей и теории нечетких множеств; подходом к вознаграждению участников работы распределенного реестра, основанном на системе социального кредитования;

Предлагаемое решение может найти применение как на уровне организаций различных форм собственности, так и на уровне государства в целом.

Использование предложенного метода как дополнительной меры[1] защиты персональных данных, позволит повысить: доступность и целостность персональных данных за счет отсутствия подверженного отказу единого центра обработки данных; доступность персональных данных за счет распределенной обработки данных и объединения разнородных персональных данных в единую взаимосвязанную систему; достоверность персональных данных за счет регулярного хеширования данных с использованием хорошо изученных криптографических методов, а также автоматизированной оценки рисков недостоверности обрабатываемых персональных данных.

Метод позволит обеспечить высокую степень конфиденциальности за счет применения средств криптографической защиты информации по отношению ко всем обрабатываемым персональным данным; снизить расходы на обеспечение обработки персональных данных за счет привлечения вычислительных ресурсов граждан – субъектов персональных данных.

Кроме того, построение единой взаимосвязанной системы, дающей объективную оценку индивидуму, создаст предпосылки для культивирования общей добропорядочности граждан, формирования атмосферы доверия, гармонизации общественных отношений и повышения безопасности субъекта Российской Федерации в целом.

Литература

1. Гибадуллин Р.Ф., Зиннатов А.М., Перухин М.Ю., Гайнуллин Р.Н. Реализация механизма репликации в СУБД Postgre SQL // Вестник технологического университета. 2017. Т. 20. № 24. С. 100–101.
2. Лазарева Н.Б. Анализ данных мониторинга репликации СУБД MySQL // Ученые заметки ТОГУ. 2017. Т. 8. № 3. С. 220–222.

References

1. Gibadullin R., Perukhin M., Zinnatov A., Gainullin R. Implementation of the replication mechanism in Postgre SQL DBMS. *Bulletin of the Technological University*, 2017, vol. 20, no. 24, pp. 100–101. (in Russian)
2. Lazareva N.B. Analysis of data monitoring replication of MySQL DBMS. *Uchenye zametki TOGU*, 2017, vol. 8, no. 3, pp. 220–222. (in Russian)

[1] В отношение РРПДн как ИСПДн должны применяться меры защиты информации, установленные ФСТЭК России и ФСБ России

3. Кирносов В.Ю., Куржангулов Н.М. Сравнительный анализ механизмов репликаций данных в различных СУБД // Фундаментальные и прикладные исследования в современном мире. 2017. № 18-1. С. 84–91.
4. Мыльников В.А., Елина Т.Н. Повышение оперативности и надежности облачной инфраструктуры на базе распределенной файловой системы // Актуальные вопросы естествознания. Сборник материалов III Всероссийской научно-практической конференции с международным участием, 5 апреля 2018 г. Иваново: ФГБОУ ВО «Ивановская пожарно-спасательная академия Государственной противопожарной службы МЧС России», 2018. С. 266–268.
5. Козин И.С. Метод разработки автоматизированной системы управления информационной безопасностью региональной информационной системы // Сборник трудов «Региональная информатика и информационная безопасность. Выпуск 3». СПб.: СПОИСУ, 2017. С. 284–290.
6. Козин И.С. Метод разработки автоматизированной системы управления информационной безопасностью распределённой информационной системы // Информация и космос. 2018. № 3. С. 80–88.
7. Беззатеев С.В., Волошина Н.В., Санкин П.С. Методика расчета надежности сложных систем, учитывающая угрозы информационной безопасности // Информационно-управляющие системы. 2014. № 3. С. 78–83.
8. Беззатеев С.В., Волошина Н.В. Маскирующее сжатие на основе модели взвешенной структуры изображения // Информационно-управляющие системы. 2017. № 6. С. 88–95. doi: 10.15217/issn1684-8853.2017.6.88
9. Ринчинов А.Б. Перспективы внедрения системы социального кредита в Китае, опыт Ханчжоу // Социально-политическая ситуация накануне XIX съезда КПК: Материалы ежегодной научной конференции Центра политических исследований и прогнозов ИДВ РАН, 15-17 марта 2017 г. М.: Институт Дальнего Востока РАН, 2017. С. 348–357.
10. Улезло Д.С., Кадан А.М. Методы машинного обучения в решении задач информационной безопасности // Proc. 3rd International Conference Intelligent Technologies for Information Processing and Management (ITIPM'2015). Vol. 1. Ufa: USATU, 2015. С. 41–44.
11. 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
12. Колмогоров А.Н. О представлении непрерывных функций нескольких переменных в виде суперпозиций непрерывных функций одного переменного и сложения // Доклады АН СССР. 1957. Т. 114. № 5. С. 953–956.
13. Rumelhart D.E., Hinton G.E., Williams R.J. Learning representations by back-propagating errors // Nature. 1986. V. 323. P. 533–536. doi: 10.1038/323533a0
14. Hertz J., Krogh A., Palmer R. Introduction to the theory of neural computation. Addison Wesley, Redwood City, 1991. 327 p.

3. Kirnosov V.Yu., Kurzhangulov N.M. Comparative analysis of data replication mechanisms in various DBMS. *Funamental'nye i prikladnye issledovaniya v sovremennom mire*, 2017, no. 18-1, pp. 84–91. (in Russian)
4. Mylnikov V.A., Yelina T.N. Increase the efficiency and reliability of the cloud infrastructure based on a distributed file system. *Actual Problems of Natural Science: Proc. 3rd All-Russian Conf. Ivanovo, Russia*, 2018, pp. 266–268. (in Russian)
5. Kozin I. Method of development of security information and event management computer- aided systemof regional information system. *Proc. Regional Informatics and Information Security. The Issue No 3*, 2017, pp. 284–290. (in Russian)
6. Kozin I. Method of development of an automated information security control system for distributed applications. *Information and Space*, 2018, no. 3, pp. 80–88. (in Russian)
7. Bezzateev S.V., Voloshina N.V., Sankin P.S. Safety analysis methodology of complex systems taking into account the threats to information security. *Information and Control Systems*, 2014, no. 3, pp. 78–83. (in Russian)
8. Bezzateev S.V., Voloshina N.V. Masking compression based on weighted image structure model. *Information and Control Systems*, 2017, no. 6, pp. 88–95. (in Russian). doi: 10.15217/issn1684-8853.2017.6.88
9. Rinchinov A.B. The prospects of social credit system project in China, Hangzhou experience. *Proc. annual scientific conf. «Socio-Political Situation in China before 19th Congress of the CPC»*. Moscow, Center on Political Research and Prognoses Institute of the Far Eastern Studies, 2017, pp. 348–357. (in Russian)
10. Ulezlo D.S., Kadan A.M. Machine learning methods in solving information security problems. *Proc. 3rd International Conference Intelligent Technologies for Information Processing and Management (ITIPM'2015)*. Vol. 1. Ufa, USATU, 2015, pp. 41–44. (in Russian)
11. Kozin I.S. Providing personal data protection in an information system based on user behavior analytics. *Information and Control Systems*, 2018, no. 3, pp. 69–78. (in Russian). doi: 10.15217/issn1684-8853.2018.3.69
12. Kolmogorov A.N. On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. *Dokl. Akad. Nauk SSSR*, 1957, vol. 114, no. 5, pp. 953–956. (in Russian)
13. Rumelhart D.E., Hinton G.E., Williams R.J. Learning representations by back-propagating errors. *Nature*, 1986, vol. 323, pp. 533–536. doi: 10.1038/323533a0
14. Hertz J., Krogh A., Palmer R. *Introduction to the theory of neural computation*. Redwood City, Addison Wesley, 1991, 327 p.

Авторы

Козин Иван Сергеевич — начальник отдела защиты информации, АО «Кронштадт Технологии», Санкт-Петербург, 199178, Российская Федерация, ORCID ID: 0000-0003-3439-9702, ivan.kozin@krontech.ru

Authors

Ivan S. Kozin — Information security department head, JSC Kronshtadt Technologies, Saint Petersburg, 199178, Russian Federation, ORCID ID: 0000-0003-3439-9702, ivan.kozin@krontech.ru