

Научная статья

DOI 10.66424/2071-8217-2026-3-10

УДК 004.056.53

АЛГОРИТМ ОБЕЗЛИЧИВАНИЯ ПЕРСОНАЛЬНЫХ ДАННЫХ ПРИ ПЕРЕДАЧЕ В ТЕСТОВЫЕ СРЕДЫ РАЗРАБОТКИ В ГОСУДАРСТВЕННЫХ ОРГАНАХ И ОРГАНИЗАЦИЯХ

А. Д. Барышникова, И. М. Морозов*

Российский государственный университет нефти и газа
(национальный исследовательский университет) имени И. М. Губкина, Москва, Россия

✉ *morozowilui@gmail.com

ДЛЯ ЦИТИРОВАНИЯ

Барышникова А. Д., Морозов И. М.
Алгоритм обезличивания
персональных данных при передаче
в тестовые среды разработки
в государственных органах
и организациях // Проблемы
информационной безопасности.
Компьютерные системы.
2026. № 3. С. 131–144.
DOI: 10.66424/2071-8217-2026-3-10

ПОСТУПИЛА 25.05.2026

ПРИНЯТА 02.07.2026

ОПУБЛИКОВАНА 13.10.2026

© Барышникова А. Д., Морозов И. М.
Издатель: Санкт-Петербургский
политехнический университет
Петра Великого

АННОТАЦИЯ

Цифровая трансформация государственного управления сопровождается интенсивным развитием автоматизированных информационных систем, обрабатывающих значительные массивы персональных данных граждан. Практика подготовки тестовых сред путем копирования записей из эксплуатационных баз данных создает серьезные риски несанкционированного распространения персональных данных прежде всего со стороны внутренних нарушителей, обладающих легитимным доступом к информационным ресурсам. Анализируется проблема защиты персональных данных при их передаче в тестовые среды разработки в государственных органах и организациях. Проведен анализ типовой организационной структуры и информационных потоков, формализована модель внутреннего нарушителя, рассмотрена целевая угроза УБИ.096 («Копирование») согласно базовому документу угроз ФСТЭК России. Предложен алгоритм необратимого обезличивания, сочетающий детерминированное хеширование с уникальной солью на основе алгоритма Кескак-256 и контекстно-зависимое маскирование. Алгоритм соответствует требованиям федерального закона № 152-ФЗ «О персональных данных», № 187-ФЗ «О безопасности критической информационной инфраструктуры», ГОСТ Р 57700.34–2017 и базового документа угроз ФСТЭК России. Показана интегрируемость решения в CI/CD-конвейеры без снижения гибкости разработки. Результаты демонстрируют применимость предложенного подхода как в образовательных, так и других государственных организациях, эксплуатирующих информационные системы с массивами структурированных персональных данных.

КЛЮЧЕВЫЕ СЛОВА

Персональные данные, обезличивание, тестовая среда, информационная безопасность, внутренний нарушитель, Кескак-256, CI/CD

Original article

DOI 10.66424/2071-8217-2026-3-10

PERSONAL DATA DEPERSONALIZATION ALGORITHM FOR TRANSFER TO TEST DEVELOPMENT ENVIRONMENTS IN GOVERNMENT BODIES AND ORGANIZATIONS

A. D. Baryshnikova, I. M. Morozov*

National University of Oil and Gas “Gubkin University”, Moscow, Russia

✉ *morozowilui@gmail.com

FOR CITATION

Baryshnikova A. D., Morozov I. M.
Personal data depersonalization
algorithm for transfer to test
development environments in
government bodies and organizations.
Problems of information security.
Computer systems.
2026. No. 3, pp. 131–144.
DOI: 10.66424/2071-8217-2026-3-10
(In Russian)

RECEIVED 25.05.2026

ACCEPTED 02.07.2026

PUBLICATION 13.10.2026

ABSTRACT

The digital transformation of public administration is accompanied by the intensive development of automated information systems that process significant amounts of personal data of citizens. The practice of preparing test environments by copying records from operational databases creates serious risks of unauthorized dissemination of personal data, primarily from internal intruders with legitimate access to information resources. The article analyzes the problem of personal data protection when they are transferred to test development environments in government agencies and organizations. The analysis of the typical organizational structure and information flows is carried out, the model of the internal violator is formalized, the target threat of homicide is considered.⁰⁹⁶ (“Copying”) according to the Basic Threat Document of the FSTEC of Russia. An irreversible deconstruction algorithm is proposed that combines deterministic hashing with a unique salt based on the Keccak-256 algorithm and context-sensitive masking. The algorithm complies with the requirements of Federal Law No. 152-FZ “On Personal Data”, Federal Law No. 187-FZ “On the Security of Critical Information Infrastructure”, GOST R57700.34–2017 and the Basic Threat Document of the FSTEC of Russia. The integrability of the solution into CI/CD pipelines is shown without reducing the flexibility of development. The results demonstrate the applicability of the proposed approach both in educational and other government organizations operating information systems with arrays of structured personal data.

KEYWORDS

Personal data, depersonalization, test environment, information security, internal intruder, Keccak-256, CI/CD

1. ВВЕДЕНИЕ

В условиях масштабной цифровой трансформации государственного управления органы исполнительной власти, подведомственные учреждения и операторы критической информационной инфраструктуры переходят на автоматизированные информационные системы, обрабатывающие значительные массивы персональных данных (ПДн) граждан. Регулярная модернизация таких систем – внедрение порталов государственных услуг, систем электронного документооборота, платформ дистанционного взаимодействия – неизбежно предполагает создание изолированных тестовых сред для отладки программного кода, проверки совместимости компонентов и проведения нагрузочного тестирования [1, 2].

На практике подготовка тестовых данных зачастую осуществляется путем непосредственного копирования записей из эксплуатационных (продакшн) баз данных. Тестовые среды, как правило, не под-

падают под те же требования защиты, что и основные информационные системы: они могут быть развернуты без интеграции в централизованную систему мониторинга, не подключены к средствам предотвращения утечек данных (DLP), а доступ к ним предоставляется по упрощенной схеме. В таких условиях реальные персональные данные – включая ФИО, СНИЛС, ИНН, паспортные реквизиты, адреса – оказываются в окружении с ослабленными мерами защиты, доступном широкому кругу разработчиков, тестировщиков и внешних подрядчиков¹ [3].

Согласно федеральному закону от 27.07.2006 № 152-ФЗ «О персональных данных»², оператор обязан применять

¹ Базовый документ угроз безопасности персональных данных при их обработке в информационных системах персональных данных (ИСПДн). М. : ФСТЭК России, 2023. 45 с

² Федеральный закон от 27.07.2006 № 152-ФЗ (ред. от 2024 г.) «О персональных данных» // Собрание законодательства РФ. 2006. № 31 (ч. 1). Ст. 3451.

меры по защите ПДн, включая их обезличивание при использовании в целях, не связанных с достижением целей обработки. Федеральный закон от 26.07.2017 № 187-ФЗ «О безопасности критической информационной инфраструктуры Российской Федерации»³ обязывает субъектов критической информационной инфраструктуры минимизировать риски несанкционированного распространения информации даже в непроизводственных контурах. Передача данных в тестовую среду явно относится к случаям использования ПДн вне целей первоначальной обработки, однако в большинстве государственных органов и организаций автоматизированные процедуры обезличивания по-прежнему отсутствуют.

Особую остроту проблема приобретает в контексте угрозы, формализованной в базовом документе угроз безопасности персональных данных ФСТЭК России как УБИ.096: «Несанкционированное получение (копирование) персональных данных сотрудником, имеющим легитимный доступ к информационной системе». Внутренний нарушитель – сотрудник ИТ-подразделения с расширенными правами доступа и знанием архитектуры баз данных – способен извлечь массив ПДн под предлогом технической необходимости, не нарушая штатного функционирования системы. Это обстоятельство принципиально отличает данную угрозу от внешних атак и обуславливает необходимость защиты на уровне самих данных, а не только на уровне контроля доступа [1, 3, 4].

Существующие отечественные и зарубежные решения в области обезличивания данных – статистические методы (k -анонимность, l -разнообразие, t -близость), криптографические подходы (SHA-256, AES-256), а также специализированные программные платформы (ARX, Microsoft Presidio, Delphix) – имеют существенные ограничения применительно к задачам

государственного сектора России: не поддерживают структуры русскоязычных ПДн, являются обратимыми, требуют коммерческого лицензирования либо облачных зависимостей, несовместимых с требованиями локализации данных по ФЗ-152⁴ [1]. Таким образом, в исследовательском пространстве сформировался выраженный пробел: отсутствует легковесное, необратимое, форматосохраняющее решение для структурированных ПДн, интегрируемое в CI/CD-конвейеры государственных органов и организаций.

2. ПОСТАНОВКА ЗАДАЧИ

Выявленное противоречие между нормативными требованиями к защите ПДн в тестовых средах и отсутствием практически реализуемых, соответствующих российскому законодательству инструментов их обезличивания, определяет проблему исследования.

Объектом исследования является процесс передачи структурированных ПДн из эксплуатационной среды в тестовую в информационных системах государственных органов и организаций. Предмет исследования – алгоритмические методы необратимого обезличивания ПДн, обеспечивающие сохранение семантической целостности и функциональной пригодности данных для целей тестирования.

Целью работы является разработка алгоритма необратимого обезличивания структурированных ПДн, нейтрализующего угрозу УБИ.096 («Копирование») при сохранении функциональной пригодности данных для тестирования и соответствии требованиям ФЗ-152, ФЗ-187 и ГОСТ Р 57700.34–2017⁵.

³ Федеральный закон от 26.07.2017 № 187-ФЗ «О безопасности критической информационной инфраструктуры Российской Федерации» // Собрание законодательства РФ. 2017. № 31 (ч. I). Ст. 4746.

⁴ NIST Special Publication 800–188. De-Identifying Government Datasets. Gaithersburg, MD: National Institute of Standards and Technology, 2023. 67 p. DOI: 10.6028/NIST.SP.800-188.

⁵ ГОСТ Р 57700.34–2017. Технологии защиты информации. Обезличивание персональных данных. Термины и определения. М.: Стандартинформ, 2017. 12 с.

Гипотеза исследования: применение комбинированного алгоритма необратимого обезличивания, основанного на детерминированном хешировании с уникальной солью (Кессак-256) и контекстно-зависимом маскировании, позволяет исключить возможность идентификации субъекта ПДн при сохранении функциональной пригодности данных для тестирования, что соответствует требованиям ФЗ-152, ФЗ-187 и ГОСТ Р 57700.34–2017. Методологическую основу составляет оценка защищенности по модели STRIDE [5], адаптированная к специфике внутренних угроз в государственных информационных системах.

Для достижения поставленной цели сформулированы следующие задачи: провести анализ типовой организационной структуры государственной организации и информационных потоков; разработать формальную модель внутреннего нарушителя; описать целевую угрозу «Копирование»; предложить алгоритм обезличивания и оценить его эффективность по критериям невозможности идентификации и сохранения функциональности данных.

3. АНАЛИЗ ОБЪЕКТА ЗАЩИТЫ

Обзор существующих подходов. Проблема защиты ПДн при подготовке тестовых сред исследуется в двух магистральных направлениях: статистической анонимизации и криптографическом преобразовании данных.

Методы на основе *k*-анонимности [6, 7] обеспечивают защиту от атак связывания, однако требуют значительных вычислительных ресурсов и приводят к потере информации, что снижает пригодность данных для тестирования бизнес-логики. Подходы, основанные на дифференциальной приватности⁶, хотя и предоставляют математические гарантии конфиденциальности, вводят искусственный шум,

делающий их неприменимыми в сценариях, где требуется точное воспроизведение форматов и диапазонов, например, при валидации e-mail или ИНН.

Криптографические методы, прежде всего хеширование по алгоритмам SHA-256 и Кессак-256, обеспечивают необратимость, соответствующую требованиям федерального закона № 152-ФЗ⁷. Вместе с тем прямое применение хеш-функций нарушает сохранение формата – хеш от адреса электронной почты не является валидным e-mail-адресом, что затрудняет тестирование интерфейсов и валидаторов. Метод оценки безопасности на основе методологии STRIDE, примененный к подобным системам [5], позволяет идентифицировать угрозы раскрытия информации (Information Disclosure) как наиболее актуальный класс рисков именно в сценарии передачи данных в тестовые среды.

Существующие программные решения (ARX Data Anonymization Tool, Microsoft Presidio, Delphix) имеют существенные ограничения для госсектора России: отсутствие поддержки русскоязычных идентификаторов (СНИЛС, ИНН), обратимость преобразований, требование коммерческого лицензирования и облачных зависимостей, несовместимых с требованиями локализации данных по ФЗ-152⁸ [1, 8–10]. Системы класса Webiomed.Anonymizer демонстрируют высокие показатели F1-меры (>0,92) для медицинских текстов на русском языке, однако ориентированы на неструктурированные данные и не адаптированы к задачам подготовки тестовых баз данных [8].

Отдельную группу образуют отечественные средства маскирования, встроенные в системы управления базами данных. Расширение pgpro_anonymizer (Postgres Pro Enterprise, доступно с версии 15.1.1) реализует статическое и динамическое маскирование, а также формирование обезличенного дампа для

⁶ NIST Special Publication 800-188. De-Identifying Government Datasets.

⁷ ГОСТ Р 57700.34–2017. Технологии защиты информации. Обезличивание персональных данных. Термины и определения.

⁸ NIST Special Publication 800-188. De-Identifying Government Datasets.

сред тестирования и разработки на основе декларативных правил, задаваемых средствами DDL. Инструмент `pg_anon` («Тантор Лабс») обеспечивает автоматизированную анонимизацию данных PostgreSQL по словарям правил с режимами создания и восстановления дампов. Указанные решения применимы и активно развиваются, однако обладают двумя принципиальными для рассматриваемой задачи ограничениями. Во-первых, они привязаны к конкретной СУБД и реализуют преобразование внутри ее ядра, тогда как для государственных информационных систем характерна гетерогенность хранилищ. Во-вторых, преобладающие в них стратегии (фальсификация, перестановка, динамическое маскирование) не гарантируют необратимости в смысле ГОСТ Р 57700.34–2017: динамическое маскирование сохраняет исходные данные в базе, а фальсификация и псевдонимизация по своей природе обратимы. Таким образом, встроенные средства СУБД решают близкую, но нетождественную задачу, что обуславливает потребность в платформенно-независимом необратимом преобразовании на этапе подготовки тестовых данных.

Анализ отечественных работ в области автоматического обезличивания конфиденциальной информации [11] показывает устойчивую тенденцию к комбинированию детерминированного хеширования с контекстно-зависимым маскированием для структурированных данных. Подходы, предложенные в работах по обфускации данных в системах хранения и обработки больших объемов информации, подтверждают применимость гибридных алгоритмов в условиях распределенных информационных систем государственного сектора. Аналогичные выводы формулируются в обзоре методов анонимизации для задач машинного обучения [6].

Объект защиты и информационные потоки. Объектом защиты в рамках работы выступает информационная система государственной организации, осуществляющая обработку, хранение и передачу ПДн граждан. В типовой государственной

организации обработкой ПДн занимается несколько функциональных подразделений. Отдел кадров обрабатывает данные сотрудников: ФИО, паспортные данные, СНИЛС, ИНН, сведения о трудовой деятельности. Служба поддержки граждан работает с обращениями, заявлениями и анкетами, содержащими контактные и идентификационные данные. ИТ-департамент отвечает за разработку, сопровождение и тестирование информационных систем, обладая наиболее широкими правами доступа ко всем массивам данных. Финансово-бухгалтерский отдел обрабатывает данные, связанные с выплатами и налоговыми вычетами [1, 2].

Информационные потоки между подразделениями осуществляются через централизованные базы данных, API-интерфейсы между внутренними и внешними сервисами, а также файловые выгрузки в форматах CSV, Excel, XML для задач миграции и тестирования [2, 9].

В государственных органах и организациях обрабатываются ПДн различных категорий по постановлению Правительства РФ от 01.11.2012 № 1119. Общедоступные данные включают ФИО, паспортные сведения, контактную информацию, адрес, образование и профессию. Специальные категории охватывают сведения о национальности, состоянии здоровья, религиозных убеждениях и судимостях. Биометрические данные – фотоизображения, отпечатки пальцев, биометрические параметры – обрабатываются в ряде ИС, связанных с системами контроля и управления доступом^{9,10}.

Принципиально важно, что наибольшую угрозу представляет не столько наличие прямых идентификаторов, сколько их комбинация с косвенными признаками. Анализ успешно проведенных атак деанонимизации демонстрирует тревожную тенденцию: для уникальной идентификации личности в большинстве случаев

⁹ Федеральный закон от 27.07.2006 № 152-ФЗ (ред. от 2024 г.) «О персональных данных» // Собрание законодательства РФ.

¹⁰ ГОСТ Р 57700.34–2017. Технологии защиты информации. Обезличивание персональных данных. Термины и определения.

достаточно двух-трех квазиидентификаторов [7, 11]. Это обстоятельство определяет ключевое требование к разрабатываемым алгоритмам: они должны быть устойчивы к атакам с использованием вспомогательной информации из открытых источников (linkage attacks).

4. МОДЕЛЬ УГРОЗ И НАРУШИТЕЛЯ

Обоснование выбора целевой угрозы. При построении модели угроз использованы методические документы ФСТЭК России, в частности базовый документ угроз безопасности ПДн, а также методология оценки безопасности информационных систем STRIDE, примененная к задаче выявления угроз в сценарии передачи структурированных данных [5]. Среди множества потенциальных угроз – блокирование, модификация, уничтожение, несанкционированный доступ – в контексте подготовки тестовых сред наибольшую практическую значимость имеет угроза «Копирование», формализованная как УБИ.096.

Выбор данной угрозы в качестве целевой обоснован следующими факторами:

- угроза реализуется без использования технических уязвимостей или сложных методов компрометации – исключительно посредством штатных инструментов и легитимных прав доступа;
- копирование данных не вызывает сбоев в работе информационной системы, что существенно затрудняет его обнаружение средствами мониторинга и аудита;
- последствия реализации угрозы имеют высокую степень тяжести – несанкционированное распространение ПДн влечет административную ответственность по ст. 13.11 КоАП РФ, штрафы Роскомнадзора и репутационные потери;
- угроза может быть эффективно нейтрализована алгоритмическим преобразованием данных без кардинального изменения архитектуры информационной системы или усиления контроля доступа [1, 3, 4].

Классификация нарушителей. В контексте защиты ПДн в государственных ин-

формационных системах нарушители подразделяются на внутренних и внешних [12]. Внутренние нарушители – лица, имеющие легитимный доступ к информационным ресурсам в силу должностных обязанностей, – подразделяются на три категории: сотрудников ИТ-подразделений с расширенными правами доступа; сотрудников функциональных подразделений (кадры, бухгалтерия, служба поддержки), имеющих нормативно ограниченный доступ к ПДн; руководящий состав с административными полномочиями и возможностью обхода стандартных процедур контроля. Внешние нарушители могут получить несанкционированный доступ через эксплуатацию уязвимостей, социальную инженерию или использование скомпрометированных учетных данных.

В случае тестовых сред основной вектор угрозы смещается в сторону внутренних нарушителей: именно ИТ-специалисты иницируют процесс выгрузки данных и контролируют их содержание. Наиболее опасным является сочетание высокого уровня привилегий и корыстной мотивации: такой нарушитель способен скрытно извлечь большой объем ПДн, маскируя действия под служебные задачи [1, 3, 4].

Формальная модель нарушителя. Для целей проектирования защитной меры формализуется модель внутреннего нарушителя – сотрудника ИТ-подразделения государственной организации, участвующего в разработке и тестировании информационных систем.

Цель нарушителя – получение массива ПДн субъектов с возможностью последующего использования в личных целях, передачи третьим лицам или коммерческой эксплуатации. Нарушитель стремится минимизировать риск обнаружения, маскируя свои действия под штатные операции технического обслуживания системы.

Портрет нарушителя: сотрудник ИТ-подразделения, имеющий легитимный доступ к продакшн-базам данных, права на выполнение операций экспорта и резервного копирования, знания языков запросов (SQL) и скриптовых языков (Python, Bash, PowerShell), доступ к тестовым серверам и возможность работать с данными

вне прямого контроля системы аудита. Мотивация нарушителя может определяться финансовой выгодой, недовольством условиями труда или желанием использовать данные в личных проектах.

Направления атаки: извлечение данных из продакшн-среды под предлогом подготовки тестового набора; передача необработанной копии базы данных в тестовую среду без применения процедур обезличивания; сохранение локальной копии на персональном устройстве, внешнем носителе или в облачном хранилище; повторное использование данных в сторонних проектах после увольнения.

Инструменты: стандартные утилиты СУБД (pg_dump, mysqldump, sqlcmd), GUI-клиенты (DBeaver, SSMS), скрипты автоматизации на Python, PowerShell и Bash, средства удаленного доступа (SSH, RDP), локальные средства хранения (USB-накопители, облачные сервисы).

Принимаются следующие допущения: нарушитель действует в одиночку; не использует вредоносное ПО или технические эксплойты – только легитимные инструменты; его целью является исключительно копирование данных, а не их модификация или уничтожение; нарушитель не имеет доступа к криптографическим ключам или средствам аппаратной защиты.

5. АЛГОРИТМ НЕОБРАТИМОГО ОБЕЗЛИЧИВАНИЯ

Требования к алгоритму. На основе проведенного анализа нормативной базы и модели угроз к разрабатываемому алгоритму предъявляются следующие требования^{11,12,13}:

¹¹ Федеральный закон от 27.07.2006 № 152-ФЗ (ред. от 2024 г.) «О персональных данных» // Собрание законодательства РФ.

¹² ГОСТ Р 57700.34–2017. Технологии защиты информации. Обезличивание персональных данных. Термины и определения.

¹³ Методические рекомендации по обезличиванию персональных данных. М.: Роскомнадзор, 2021. 28 с.

- необратимость – результат преобразования не позволяет восстановить исходные данные никакими средствами, включая полный доступ к тестовой базе, что соответствует определению обезличивания в ГОСТ Р 57700.34–2017;

- форматосохранение – преобразованные данные сохраняют синтаксическую структуру исходных полей, обеспечивая корректную работу интерфейсов, валидаторов и бизнес-логики тестируемых систем;

- детерминированность – одинаковые входные значения в пределах одного сеанса обезличивания преобразуются в одинаковые выходные значения, что обеспечивает сохранение ссылочной целостности базы данных;

- устойчивость к атакам реидентификации – алгоритм должен противостоять linkage attacks на основе комбинации квазиидентификаторов [6, 7];

- интегрируемость в CI/CD – реализация в форме конфигурируемого модуля предобработки данных без нарушения гибкости процессов разработки [2].

Структура алгоритма. Предлагаемый алгоритм реализует двухкомпонентный подход к обезличиванию структурированных ПДн. Первый компонент (детерминированное хеширование с уникальной солью) применяется к прямым идентификаторам: ФИО, ИНН, СНИЛС, паспортным данным, адресам электронной почты. Второй компонент (контекстно-зависимое маскирование) применяется к полям, для которых критично сохранение формата и статистических свойств: датам рождения, адресам, номерам телефонов.

Для первого компонента используется алгоритм Кескак-256 (SHA-3) – криптографически стойкая хеш-функция, обеспечивающая необратимость преобразования. Для каждого атрибута генерируется уникальная соль, что исключает возможность использования радужных таблиц и предотвращает сопоставление данных между различными тестовыми средами. Детерминированность в пределах одного сеанса обезличивания достигается фиксацией соли в конфигурационном профиле, который хранится отдельно от тестовой базы

данных и уничтожается после завершения сеанса¹⁴ [2].

Второй компонент (контекстно-зависимое маскирование) обеспечивает сохранение формата (Format-Preserving Masking) для полей, семантика которых критична для тестирования. Дата рождения заменяется псевдослучайной датой в том же возрастном диапазоне (± 5 лет) для корректной работы логики возрастных ограничений. Адрес заменяется валидным адресом того же региона и города из синтетического репозитория, обеспечивая корректность работы геолокационных сервисов и валидаторов ФИАС. Номер телефона генерируется в том же коде оператора и регионе для сохранения работоспособности систем SMS-уведомлений в тестовом режиме.

Последовательность выполнения алгоритма. Алгоритм обезличивания выполняется в следующей последовательности шагов:

1. Загрузка конфигурационного профиля обезличивания. Профиль задает соответствие между столбцами базы данных и методами преобразования (хеширование или форматосохраняющее маскирование), а также уникальные соли для каждого атрибута. Профиль хранится в защищенном хранилище и недоступен разработчикам тестовой среды.

2. Классификация полей по типу преобразования. Поля разделяются на три группы в зависимости от энтропии входного пространства. Высокоэнтропийные прямые идентификаторы (полное написание ФИО, адреса электронной почты, текстовые реквизиты) направляются на хеширование с уникальной солью. Низкоэнтропийные структурированные идентификаторы с малым пространством допустимых значений (СНИЛС, ИНН, номер телефона, номер студенческого билета) направляются на необратимую форматосохраняющую генерацию: для них синтезируется новое валидное значение с

корректной контрольной суммой, не являющееся функцией исходного и не сопровождаемое хранением соответствия, что исключает атаку полным перебором по ограниченному пространству входов. Косвенные идентификаторы (дата рождения, адрес) направляются на форматосохраняющее маскирование.

3. Хеширование высокоэнтропийных прямых идентификаторов. Для каждого поля вычисляется значение Кессак-256(соль_поля || значение). Усечение хеша, снижающее прообразную стойкость, не применяется: синтаксическая корректность обеспечивается формой шаблона поля, а не сокращением длины хеша (например, для e-mail формируется псевдослучайный адрес вида hash_N@domain.test). Низкоэнтропийные структурированные идентификаторы (СНИЛС, ИНН, номер телефона) на данном шаге не обрабатываются, для них применяется необратимая форматосохраняющая генерация (шаг 2), при которой перебор по пространству входов невозможен, поскольку выходное значение не является функцией исходного.

4. Форматосохраняющее маскирование косвенных идентификаторов. Применяются правила генерации псевдозначений с учетом семантических ограничений поля (возрастной диапазон, региональный классификатор, код оператора связи).

5. Верификация ссылочной целостности. Проверяется корректность внешних ключей: детерминированность хеш-функции гарантирует, что связанные записи в разных таблицах сохраняют согласованность.

6. Оценка риска реидентификации. Проводится анализ остаточной комбинаторной уязвимости квази-идентификаторов методом оценки, аналогичным модели k -анонимности [6], с учетом методик контроля качества обезличивания [13, 14]. При выявлении уязвимых сочетаний применяется дополнительное обобщение соответствующих атрибутов.

7. Передача обезличенного набора данных в тестовую среду и уничтожение конфигурационного профиля соли.

¹⁴ ГОСТ Р 57700.34–2017. Технологии защиты информации. Обезличивание персональных данных. Термины и определения.

Интеграция в CI/CD-конвейер. Алгоритм реализован в виде конфигулируемого модуля на языке Python, интегрируемого в CI/CD-конвейер на этапе подготовки тестовых данных. Модуль предоставляет интерфейс командной строки и может быть вызван из скрипта GitLab CI/CD, Jenkins Pipeline или аналогичного инструмента оркестрации. Параметры обезличивания задаются через YAML-профиль, хранящийся в защищенном репозитории секретов (HashiCorp Vault или аналог). Это исключает возможность несанкционированного изменения профиля разработчиками, не обладающими соответствующими правами доступа [2].

Время работы алгоритма складывается из трех составляющих: криптографического преобразования, форматосохраняющего маскирования и оценки остаточного риска реидентификации. Криптографическая составляющая является линейной по числу записей и не образует узкого места: преобразование Кессак-256 выполняется со скоростью порядка 10^6 записей в секунду на одном ядре стандартного серверного оборудования (хеширование таблицы из 100 000 записей занимает доли секунды). Основное время приходится на форматосохраняющее маскирование (генерация синтетических адресов, телефонов) и анализ комбинаций квази-идентификаторов – именно эти этапы определяют наблюдаемую длительность на реальных данных. Ядро алгоритма обрабатывает записи независимо и допускает построчное распараллеливание, а этап оценки реидентификации реализуется на выборочной (агрегатной) основе вместо попарного сравнения, что сохраняет линейный характер общей сложности и обеспечивает применимость к таблицам в миллионы записей без существенной задержки конвейера сборки.

6. ОЦЕНКА ЭФФЕКТИВНОСТИ АЛГОРИТМА

Критерии оценки. Эффективность алгоритма оценивается по двум группам критериев:

- критерии защиты конфиденциальности: невозможность восстановления исходных ПДн из обезличенных данных (необратимость по ГОСТ Р 57700.34–2017); устойчивость к атакам реидентификации на основе комбинации квази-идентификаторов; соответствие требованию об обезличивании ФЗ-152;

- критерии функциональной пригодности: сохранение ссылочной целостности базы данных; синтаксическая корректность преобразованных значений для корректной работы валидаторов; сохранение статистических свойств данных, необходимых для тестирования бизнес-логики.

Анализ соответствия требованиям регуляторов. Предлагаемый алгоритм удовлетворяет ключевым нормативным требованиям. В части ФЗ-152: применение Кессак-256 с уникальной солью обеспечивает невозможность идентификации субъекта без дополнительной информации (конфигурационного профиля соли), что соответствует определению обезличивания в ст. 3 ФЗ-152. В части ГОСТ Р 57700.34–2017: необратимость преобразования достигается криптографической стойкостью Кессак-256 (устойчивость к прообразу – не менее 2^{256} операций). В части ФЗ-187: алгоритм применим в производственных контурах критической информационной инфраструктуры, минимизируя риски несанкционированного распространения ПДн на этапах разработки и тестирования, где требования к защите данных традиционно ослаблены.

Сравнительный анализ с существующими решениями (см. таблицу) показывает, что предлагаемый алгоритм обеспечивает редкое для доступных средств сочетание необратимости, форматосохранения и полного соответствия ФЗ-152 применительно к задачам подготовки структурированных тестовых данных в государственных органах и организациях, не требуя при этом привязки к конкретной СУБД. В отличие от встроенных средств маскирования отечественных СУБД (pgpro_anonymizer, pg_anon), реализующих преимущественно обратимые стратегии в рамках одной платформы, предлагаемый подход

Сравнительный анализ методов обезличивания ПДн применительно к задачам тестирования в государственных органах и организациях

Comparative analysis of PD anonymization methods applied to testing tasks in government agencies and organizations

Метод	Необратимость	Форматосохранение	Применимость для тестирования	Соответствие ФЗ-152
К-анонимность	Нет	Частичное	Средняя	Частичное
Дифференциальная приватность	Да	Нет	Низкая	Да
Замена идентификатора (псевдонимизация)*	Нет	Высокое	Очень высокая	Нет
Хеширование (SHA-256)	Да	Нет	Средняя	Да
Маскирование данных	Нет	Высокое	Высокая	Частичное
Встроенное маскирование СУБД (pgpro_anonymizer, pg_anon)	Частичная	Высокое	Высокая	Частичное
Кескак-256 + маскирование (предлагаемый подход)	Да	Высокое	Высокая	Да

*Замена идентификатора (псевдонимизация) является обратимой и не удовлетворяет критерию необратимости по ГОСТ Р 57700.34–2017.

выполняет необратимое преобразование высокоэнтропийных идентификаторов в платформенно-независимом модуле преобработки.

Апробация на примере РГУ нефти и газа им. И. М. Губкина. В качестве иллюстративного полигона для апробации алгоритма выбрана образовательная организация, эксплуатирующая крупную ИТ-инфраструктуру с массивами ПДн студентов и сотрудников, типичную для государственного сектора.

В такой организации функционируют: образовательный портал; системы управления обучением (LMS); базы данных по успеваемости, практикам и научно-исследовательской деятельности; модули интеграции с промышленными партнерами. При разработке новых функциональных модулей ИТ-специалисты копируют фрагменты продакшн-баз в тестовые среды. Без применения обезличивания подобные действия нарушают требования ФЗ-152 и создают угрозу утечки данных, особенно с учетом того, что сведения о студентах часто связаны с допусками на объекты повышенной опасности, ме-

дицинскими осмотрами и специальными проверками [1, 8].

Применение предложенного алгоритма к тестовой выгрузке базы данных учебной части (12 000 записей студентов) показало следующие результаты. Все прямые идентификаторы (ФИО, номера студенческих билетов, e-mail-адреса корпоративного домена) необратимо преобразованы с сохранением формата: хешированные e-mail остаются валидными с доменом @test.local. Даты рождения смещены в пределах ± 5 лет с сохранением корректности возрастной логики. Ссылочная целостность базы данных полностью сохранена. Оценка остаточного риска реидентификации по методу квази-идентификаторов показала отсутствие уязвимых комбинаций при используемых параметрах обезличивания. Время обработки составило 9,4 с, что не оказало значимого влияния на продолжительность CI/CD-конвейера. Профилирование показало, что в указанном времени доля криптографического преобразования составляет менее 0,1 с, а основной вклад вносят форматосохраняющее маскирование и оценка остаточ-

ного риска реидентификации. Поскольку обе доминирующие операции линейны по числу записей и распараллеливаемы, экстраполяция на таблицы большего объема выполняется пропорционально: при сохранении выборочной схемы оценки реидентификации обработка миллиона записей укладывается в десятки секунд на одном узле и далее сокращается распараллеливанием.

7. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Полученные результаты демонстрируют, что комбинированный алгоритм, основанный на детерминированном хешировании Кессак-256 с уникальной солью и контекстно-зависимом форматосохраняющем маскировании, превосходит альтернативные подходы по совокупности критериев, критически важных для государственного сектора России: необратимость + форматосохранение + соответствие ФЗ-152 + интегрируемость в CI/CD.

Необходимо подчеркнуть принципиальное отличие предлагаемого подхода от классической замены идентификатора (псевдонимизации): обратимость последней не соответствует критерию необратимости по ГОСТ Р 57700.34–2017 и, следовательно, не является правомерным методом обезличивания по ФЗ-152. В то же время прямое хеширование без форматосохранения, хотя и обеспечивает необратимость, нарушает функциональную пригодность данных для тестирования интерфейсов и валидаторов. Предлагаемый алгоритм устраняет это противоречие за счет дифференцированного применения методов в зависимости от семантической роли атрибута. Применение методологии STRIDE [5] для верификации модели угроз подтвердило, что наиболее актуальным классом рисков в рассматриваемом сценарии является Information Disclosure (раскрытие информации), а предлагаемый алгоритм непосредственно нейтрализует данный класс угроз на уровне данных.

Ограничения работы определяются допущениями принятой модели нарушителя.

Алгоритм не рассматривает сценарии сговора нескольких сотрудников, гибридных атак с использованием вредоносного программного обеспечения, угрозы целостности и доступности данных. Алгоритм ограничен структурированными данными реляционных СУБД и не распространяется на неструктурированные форматы (документы, изображения, медицинские тексты), требующие иных подходов к обезличиванию, включая NER-модели, оптимизированные для русскоязычных текстов [8, 15].

Перспективными направлениями развития предложенного подхода являются: расширение поддержки типов данных, включая JSONB-поля PostgreSQL и неструктурированные вложения; разработка адаптивного профиля обезличивания с автоматическим определением типов ПДн на основе машинного обучения; интеграция с российскими СУБД (Postgres Pro, Arenadata); разработка инструментальных средств для аудита соответствия результатов обезличивания требованиям регуляторов в автоматическом режиме.

8. ЗАКЛЮЧЕНИЕ

Рассмотрена проблема защиты ПДн при их передаче в тестовые среды разработки в информационных системах государственных органов и организаций. Проведенный анализ показал, что практика копирования данных из продакшн-баз без предварительного обезличивания является широко распространенной, при этом нарушающей требования ФЗ-152 и создающей реальные риски реализации угрозы УБИ.096 – несанкционированного копирования ПДн внутренним нарушителем.

Формализованная модель нарушителя описывает сотрудника ИТ-подразделения с расширенными правами доступа, действующего исключительно с использованием легитимных инструментов и маскирующего копирование данных под штатные технические операции. Анализ существующих методов обезличивания выявил, что ни один из них в отдельности

не обеспечивает одновременного выполнения требований необратимости, форматосохранения и интегрируемости в CI/CD-процессы государственных органов и организаций.

Предложенный алгоритм, сочетающий детерминированное хеширование Кессак-256 с уникальной солью для прямых идентификаторов и форматосохраняющее маскирование для косвенных идентификаторов, устраняет выявленный исследовательский пробел. Алгоритм соответствует требованиям ФЗ-152, ФЗ-187 и ГОСТ Р 57700.34–2017, интегрируется в

CI/CD-конвейеры без снижения гибкости разработки и демонстрирует практическую применимость на примере образовательной организации.

Практическая значимость работы определяется тем, что предложенное решение может быть непосредственно внедрено в любой государственной организации, осуществляющей разработку и тестирование информационных систем с массивами структурированных ПДн, без приобретения коммерческих лицензий и введения облачных зависимостей, несовместимых с требованиями локализации данных.

КОНФЛИКТ ИНТЕРЕСОВ / CONFLICT OF INTERESTS

Авторы заявляют об отсутствии конфликта интересов / The authors declare no conflict of interests.

СПИСОК ИСТОЧНИКОВ

1. **Замараева О. А., Титов В. А., Кузин Д. О.** Разработка политики информационной безопасности для экономического вуза: определение информации, подлежащей защите, и построение модели злоумышленника // *Современные проблемы науки и образования*. 2014. № 3. С. 330–330. EDN: SYZOLH.
2. **Бабак Н. Г., Белорыбкин Л. Ю., Оцков Ш. А. и др.** Автоматическое обезличивание конфиденциальной информации // *Российский технологический журнал*. 2023. Т. 11. № 5. С. 7–18. DOI: 10.32362/2500-316X-2023-11-5-7-18. EDN: JMYNZQ.
3. **Melnyk A., Le Boudec J.-Y., Morozov Y. et al.** Protection of Data Transmission in Remote Monitoring Tools by Anonymization // *IntelITSIS'2023: 4th International Workshop on Intelligent Information Technologies and Systems of Information Security*, March 22–24, 2023, Khmelnytskyi, Ukraine. Khmelnytskyi, Ukraine, 2023. P. 24–37.
4. **Shamsinejad E., Rahmani A.M., Hosseinzadeh M.** A Review of Anonymization Algorithms and Methods in Big Data // *Annals of Data Science*. 2025. Vol. 12. P. 253–279. DOI: 10.1007/s40745-024-00557-w.
5. **Уймин А. Г., Морозов И. М.** Оценка безопасности WINE с использованием методологии STRIDE: математическая модель // *Современная наука: актуальные проблемы теории и практики*. Серия: Естественные и технические науки. 2023. № 6–2. С. 164–170. EDN: HYCKNP.
6. **El Ouazzani Z.** Novel Anonymization Approaches Ensuring Privacy: Application in E-Health: Doctoral dissertation. Rabat: ENSIAS, Mohammed V University, 2021.
7. **Marco-Pérez I., Pérez B., Rubio Garcia A. L., Zapata M. A.** The Many Faces of Data Deletion: On the Significance and Implications of Deleting Data // *ACM Computing Surveys*. 2026. Vol. 58. № 7. Art. 174. P. 1–36. DOI: 10.1145/3779299.
8. **Макарова Е. А., Андрейченко А. Е., Казакова М. А. и др.** Обезличивание медицинских текстовых данных с целью разработки и внедрения систем искусственного интеллекта // *Правовая информатика*. 2024. № 3. С. 96–105. EDN: XPLMNO.
9. **Murugan A.** Multi-Tenant Data Filtering in Fog Nodes for Privacy-Preserving IoT Data Processing. Dublin: National College of Ire-

- land, 2026. URL: <https://norma.ncirl.ie/id/eprint/9247> (дата обращения: 20.05.2026).
10. **Rahimli R. Sh.** Security and Privacy in Information Processing. Baku: Khazar University, 2025.
 11. **Poltavtseva M. A., Sokolov D. V., Petrov A. S., Ivanova E. K.** Optimization of Data Obfuscation in Big Data Processing and Storage Systems // Automatic Control and Computer Sciences. 2025. Vol. 59. № 8. P. 1464–1474. DOI: 10.3103/S0146411625701020.
 12. **Тихонова А. Б.** Защита персональных данных с помощью анонимизации // Материалы XIX Международной конференции по вычислительной механике и современным прикладным программным системам (ВМСППС'2015), 24–31 мая 2015 г., Алушта. Алушта: Московский авиационный институт, 2015. С. 179–180. EDN: UTPOFV.
 13. **Моисеева А. С., Кузнецова М. С.** Обезличивание данных и его контроль // Телекоммуникации и информационные технологии – реалии, возможности, перспективы: Сборник трудов XV Молодежного научного форума, 1–22 апреля 2024 г., Москва. М.: МТУСИ, 2024. С. 241–245. EDN: PMMIUU.
 14. **Иванова В. Д.** Анализ современных методов защиты персональных данных // Проблемы современной науки и ее прикладные аспекты: Сборник материалов Международной научно-практической конференции, 14 мая 2025 г., Чистополь. СПб.: ИД «Сциентиа», 2025. С. 282–288. EDN: ZZKCWI.
 15. **Столбов А. П.** Обезличивание персональных данных в здравоохранении // Врачи и информационные технологии. 2017. № 3. С. 76–91. EDN: ZGYVOT.

REFERENCES

1. **Zamaraeva O. A., Titov V. A., Kuzin D. O.** Development of policy of information security for economic higher education institution: definition of information which is subject to protection, and creation of model of the malefactor. *Sovremennye problemy nauki i obrazovaniya*. 2014. No. 3, pp. 330–330. EDN: SYZOLH. (In Russian)
2. **Babak N. G., Belorybkin L. Yu., Otsokov S. A. et al.** Automatic depersonalization of confidential information. *Russian Technological Journal*. 2023. Vol. 11. No. 5, pp. 7–18. DOI: 10.32362/2500-316X-2023-11-5-7-18. EDN: JMYNZQ. (In Russian)
3. **Melnyk A., Le Boudec J.-Y., Morozov Y. et al.** Protection of Data Transmission in Remote Monitoring Tools by Anonymization. *IntellT-SIS'2023: 4th International Workshop on Intelligent Information Technologies and Systems of Information Security*, March 22–24, 2023, Khmelnytskyi, Ukraine. Khmelnytskyi, Ukraine, 2023, pp. 24–37.
4. **Shamsinejad E., Rahmani A.M., Hosseinzadeh M.** A Review of Anonymization Algorithms and Methods in Big Data. *Annals of Data Science*. 2025. Vol. 12, pp. 253–279. DOI: 10.1007/s40745-024-00557-w.
5. **Uymin A. G., Morozov I. M.** Wine safety assessment using stride methodology: a mathematical model. *Sovremennaya nauka: aktual'nye problemy teorii i praktiki. Seriya: Estestvennye i tekhnicheskie nauki*. 2023. No. 6–2, pp. 164–170. EDN: HYCKNP. (In Russian)
6. **El Ouazzani Z.** Novel Anonymization Approaches Ensuring Privacy: Application in E-Health: Doctoral dissertation. Rabat: ENSIAS, Mohammed V University, 2021.
7. **Marco-Pérez I., Pérez B., Rubio Garcia A. L., Zapata M. A.** The Many Faces of Data Deletion: On the Significance and Implications of Deleting Data. *ACM Computing Surveys*. 2026. Vol. 58. No. 7. Art. 174, pp. 1–36. DOI: 10.1145/3779299.
8. **Makarova E. A., Andreichenko A. E., Kazakova M. A. et al.** Depersonalisation of medical text data with a view to developing and implementing artificial intelligence systems. *Pravovaya informatika*. 2024. No. 3, pp. 96–105. EDN: XPLMNO. (In Russian)
9. **Murugan A.** Multi-Tenant Data Filtering in Fog Nodes for Privacy-Preserving IoT Data Processing. Dublin: National College of Ireland, 2026. URL: <https://norma.ncirl.ie/id/eprint/9247> (accessed: 20.05.2026).
10. **Rahimli R. Sh.** Security and Privacy in Information Processing. Baku: Khazar University, 2025.
11. **Poltavtseva M. A., Sokolov D. V., Petrov A. S., Ivanova E. K.** Optimization of Data

- Obfuscation in Big Data Processing and Storage Systems. *Automatic Control and Computer Sciences*. 2025. Vol. 59. No. 8, pp. 1464–1474. DOI: 10.3103/S0146411625701020.
12. **Tikhonova A. B.** Protection of personal data using anonymization. *Materialy XIX Mezhdunarodnoj konferencii po vychislitel'noj mehanike i sovremennym prikladnym programnym sistemam (VMSPPS'2015)*, 24–31 maya 2015 g., Alushta. Alushta : Moskovskij aviacionnyj institut, 2015, pp. 179–180. EDN: UTPOFV. (In Russian)
13. **Moiseeva A. S., Kuznetsova M. S.** Data anonymization and its control. *Telekommunikacii i informacionnye tehnologii – realii, vozmozhnosti, perspektivy: Sbornik trudov XV Molodezhnogo nauchnogo foruma*, 1–22 aprelya 2024 g., Moscow. Moscow: MTUSI, 2024, pp. 241–245. EDN: PMMIUU. (In Russian)
14. **Ivanova V. D.** Analysis of modern methods of personal data protection. *Problemy sovremennoj nauki i ee prikladnye aspekty: Sbornik materialov Mezhdunarodnoj nauchno-prakticheskoy konferencii*, 14 maya 2025 g., Chis-topol'. St. Petersburg: ID «Scientia», 2025, pp. 282–288. EDN: ZZKCWI. (In Russian)
15. **Stolbov A. P.** De-identification of personal data in health care. *Medical doctor and IT*. 2017. No. 3, pp. 76–91. EDN: ZGYVOT. (In Russian)

СВЕДЕНИЯ ОБ АВТОРАХ / INFORMATION ABOUT AUTHORS

БАРЫШНИКОВА Алена Дмитриевна – лаборант, Российский государственный университет нефти и газа (национальный исследовательский университет) имени И. М. Губкина, Россия, 119991, Москва, Ленинский пр-т, д. 65
E-mail: baryshnikova1309@yandex.ru
ORCID: 0009-0002-8319-8237

BARYSHNIKOVA Alena D. – Laboratory Assistant, National University of Oil and Gas "Gubkin University", Russia, 119991, Moscow, Leninsky ave., 65

МОРОЗОВ Илья Михайлович – преподаватель, Российский государственный университет нефти и газа (национальный исследовательский университет) имени И. М. Губкина, Россия, 119991, Москва, Ленинский пр-т, д. 65
E-mail: morozowilui@gmail.com
ORCID: 0009-0002-9304-3757

MOROZOV Ilya M. – Lecturer, National University of Oil and Gas "Gubkin University", Russia, 119991, Moscow, Leninsky ave., 65