УДК 004.421.6:519.115.1

Е.Ю. Мищенко, А.Н. Соколов

Алгоритмы реализации методов обезличивания персональных данных в распределенных информационных системах

Рассматриваются различные методы защиты персональных данных (ПД), основанные на применении процедуры обезличивания, а также методика оценки эффективности обезличивания. Актуальность применения обезличивания состоит в том, что в отличие от других методов защиты (в том числе шифрования), затраты на обезличивание не зависят от количества рабочих мест, обрабатывающих ПД, что позволяет получить определенный экономический эффект при построении системы защиты. Авторы приводят описание характеристик различных методов обезличивания, которые оказывают наибольшее влияние на реализацию данных методов, а также дают обзор реализованных алгоритмов в различных сферах обработки ПД. Кроме того, приводятся результаты оценки эффективности обезличивания ПД, полученные при применении метода введения идентификаторов в сфере здравоохранения. В результате внедрения была достигнута достаточная степень защищенности ПД, а также был получен значительный экономический эффект.

Ключевые слова: вероятность идентификации, информационная система, метод введения идентификаторов, метод изменения состава или семантики, метод перемешивания, обезличивание персональных данных, степень обезличивания.

doi: 10.21293/1818-0442-2019-22-1-7-66-70

Активные методы защиты информации основаны на использовании сертифицированных средств защиты информации, интегрированных в операционную среду обработки ПД. Стоимость применения таких средств пропорциональна количеству рабочих мест. Обезличивание ПД имеет главной целью обеспечение попыток злоумышленника использовать открытую информацию во вред физическому лицу. Поскольку этот метод защиты не требует применения средств защиты, достигается еще одна цель – значительная экономия средств.

Российское законодательство так же, как и законодательство США и ЕС, предъявляет жесткие требования к защите ПД, поэтому задача их обезличивания является актуальной. Главная особенность российского законодательства заключается в том, что в соответствии с Законом «О персональных данных» [1], обезличивание не позволяет определить принадлежность ПД физическому лицу без применения дополнительной информации, т.е. наиболее важным свойством обезличенных данных является возможность их деобезличивания, а отсутствие такой возможности признается лишь как частный случай. Таким образом, в России для процесса обезличивания установлена одинаковая актуальность для прямой и обратной задачи.

Начиная с 2009 г. в России было предложено множество алгоритмов реализации процесса обезличивания, но только в 2013 г. приказом Роскомнадзора [2] были определены четыре основных метода обезличивания: введение идентификаторов, изменение состава или семантики, декомпозиция и перемешивание (таблица).

Из таблицы видно, что:

– методы идентификаторов и декомпозиции основаны на отделении идентифицирующей информации от обезличенных данных, поэтому для них общей проблемой является обеспечение связи раз-

ленных частей во время сеанса работы. При этом идентифицирующая часть недоступна для злоумышленника во время хранения, но может быть доступна во время прочих сеансов обработки (ввод, вывод). Следует отметить, что с точки зрения реализации алгоритмы этих двух методов принципиально очень близки, поэтому далее будем рассматривать их совместно;

– методы изменения состава/семантики и перемешивания основаны на скрытии расположения идентифицирующей информации в массиве обезличенных данных, поэтому для них общей проблемой является обеспечение безопасности алгоритма во время сеанса работы. При этом идентифицирующая часть является общедоступной, поэтому есть опасность вычисления алгоритма злоумышленником в любых режимах работы;

– общей, хотя и менее критичной проблемой для всех методов является выбор группы идентифицирующих атрибутов. Причиной данной проблемы является отсутствие методики количественной оценки эффективности обезличивания;

– общей проблемой, важной с точки зрения затрат, является необходимость модификации структуры базы данных и прикладного программного обеспечения для реализации метода обезличивания. В случае разработки программного обеспечения посторонней организацией требуемая модификация может оказаться экономически невыгодной и технически невозможной.

Рассмотрим варианты реализации методов обезличивания персональных данных с учетом имеющегося опыта их внедрения.

Методы обезличивания персональных данных

Основанием для анализа состояния проблемы является исследование имеющихся научных публикаций по теме обезличивания ПД в России, а также

собственные разработки авторов. Все направления исследований можно условно разделить на четыре группы по применяемым методам обезличивания.

Методы обезличивания персональных данных

Метод	Принцип работы	Секретность	Проблемы
Введение идентификаторов	Группа идентифицирующих атрибутов заменяется абстрактным идентификатором, группа хранится в отдельной таблице	Таблица перекрестных ссылок	– Выбор состава идентифицирующей группы. – Генерация идентификатора. – Обеспечение связи между таблицей и обезличенными данными
Изменение состава или семантики	Изменяется структура (количество, положение и размер полей) или изменяется значение идентифицирующих атрибутов	Алгоритм модификации	– Выбор состава идентифицирующей группы. – Генерация алгоритма модификации. – Обеспечение секретности алгоритма модификации
Декомпозиция	База данных разделяется на много частей, информация о связях хранится в отдельной таблице	Таблица связей	– Выбор состава частей. – Генерация алгоритма связывания. – Обеспечение связи между частями
Перемещение	Группа идентифицирующих атрибутов перемещается в запись другого человека	Алгоритм перемещения	– Выбор состава идентифицирующей группы. – Генерация алгоритма перемещения. – Обеспечение секретности алгоритма перемещения

1. *Алгоритмы реализации метода введения идентификаторов.* Эта группа представлена работами, часть которых рассматривает различные криптографические методы генерации идентификатора связи между таблицей перекрестных ссылок и обезличенной базой. Например, уникальный и релевантный идентификатор физического лица получен путем использования односторонней криптографической функции от атрибутов фамилии, имени, отчества и даты рождения физического лица О.А. Вишняковой и Д.Н. Лавровым [3]. Однако решение вопросов связи и безопасности сеансов обработки данных остаются за рамками этой и подобных работ.

Имеется патент на способ идентификации субъекта персональных данных с использованием в качестве идентификатора связи сим-карты, предложенный Е.С. Волокитиной [4]. Метод успешно внедрен в организациях сферы образования. Представленный алгоритм успешно решает проблему безо-

пасности во время сеансов обработки обезличенных данных, однако использование дополнительного идентификатора усложняет процесс обработки и повышает затраты. Схема реализации с использованием внешнего идентификатора представлена на рис. 1.

Рис. 1. Схема реализации с использованием SIM-карты

Для обеспечения экономической эффективности Д.Н. Ивановым и Е.Ю. Мищенко предложена полезная модель, получившая патент [5] и внедренная на предприятиях сферы здравоохранения.

Запатентованный метод в качестве идентификатора связи предлагает использовать бумажный носитель – бланк рецепта на получение лекарств. Такой идентификатор является частью технологического процесса и не вносит дополнительных сложностей и затрат при обработке данных, а сам технологический процесс предполагает размещение обезличенной базы данных локально во всех аптечных пунктах. Технологический процесс и соответствующее программное обеспечение обработки рецептов модернизированы путем автоматизированного считывания идентификаторов с бумажного носителя. Полный двусторонний обмен данными с центром обработки данных на сервере был заменен на односторонний обмен обезличенными данными от аптечного пункта к серверу. При этом на рабочих местах во время сеансов вместо чувствительных персональных данных используются абстрактные идентификаторы. Таким образом, установка средств защиты информации в аптечных пунктах не требуется, что в масштабах региона (около 100 аптечных пунктов) дает экономию более 1,5 млн руб.

А.А. Халафян и А.А. Кошкарров [6], в качестве идентификатора связи также предлагают использовать бланк рецепта на получение лекарств, но технологический процесс при этом предполагает размещение обезличенной базы данных в общем доступе в облаке с использованием каналов Интернета, что позволяет централизовать и ускорить обработку данных. Однако при этом на рабочих местах во время сеансов используются как идентификаторы, так и чувствительные атрибуты. Таким образом, все рабочие места требуют установки средств криптографической защиты и межсетевого экранирования, что не позволяет получить экономического эффекта в аптечных пунктах. Тем не менее этот метод был вне-

дрен в распределенную информационную систему сферы здравоохранения.

2. *Алгоритмы реализации метода изменения состава или семантики.* Это направление представлено работой И.Ю. Кучина [7], в которой предлагается способ кодирования идентифицирующих атрибутов на базе разработанного алгоритма. Отличительной особенностью работы является аналитическое обоснование выбора состава идентифицирующей группы и обеспечение заданной степени анонимности в обезличенной базе. Этот метод получил внедрение в сфере здравоохранения, однако вопрос обеспечения безопасности решается только при хранении персональных данных, но не для прочих режимов обработки информации.

3. *Алгоритмы реализации метода перемешивания.* Направление представлено работами, которые предлагают использование алгоритмов перемешивания, ориентированных на хранение ПД, либо на их передачу по открытым каналам связи. Например, К.О. Бондаренко и В.А. Козлов [8] представляют способ перемешивания данных внутри сегментов с последовательным перемешиванием строк и чувствительных атрибутов, причем в алгоритме используются таблицы подстановки, сгенерированные криптографическим методом гаммирования. Применение криптографии, с одной стороны, гарантирует стойкость алгоритма даже во время сеанса обработки, но, с другой стороны, усложняет процесс добавления/удаления/поиска данных и повышает затраты на защиту. Эти недостатки являются препятствием для внедрения метода.

Среди внедренных можно отметить работу М.И. Денисова и К.А. Чехонина [9]. Однако обезличенная база используется только внутри контролируемой зоны, а для передачи данных используются средства шифрования.

В.В. Воронин и Н.Л. Нехай [10] разработали алгоритм с многочисленными циклическими сдвигами в подмножествах каждого чувствительного атрибута и внедрили его в систему CRM-учета клиентов и работников предприятия по обслуживанию автотранспорта. Для хранения параметров перемешивания используются криптографические средства, что также повышает затраты на систему защиты.

4. *Прочие алгоритмы обезличивания.* Прочие направления исследований предполагают использование в основном криптографических методов (off-line либо on-line обезличивание с помощью инструментов баз данных), которые могут быть отнесены к обезличиванию с достаточной степенью условности, поскольку решают проблему невозможности идентификации физического лица по обработанным данным, но формально не входят в набор методов, установленных Роскомнадзором [2], либо используют эти методы частично. Например, работа Ю.В. Трифоновой и Р.Ф. Жаринова [11] предлагает использование встроенных криптографических средств СУБД СугртДВ. В качестве примера частичного использования метода идентификаторов можно при-

вести работу И.М. Ажмухамедова, Р.Ю. Деминой и И.В. Сафарова [12], где применяется шифрование таблицы перекрестных ссылок с последующим блокированием.

Результаты реализации методов обезличивания персональных данных

К сожалению, отсутствие общепринятой методики количественной оценки эффективности обезличивания не позволяет сравнивать результаты обезличивания данных, осуществляемого различными методами.

С целью решения проблемы предложены методики расчета эффективности обезличивания данных с использованием показателей вероятности идентификации (ВИ) и степени обезличивания для методов введения идентификаторов [13], изменения состава или семантики [14] и перемешивания [15]. На рис. 2 показаны значения ВИ для различных реквизитов, рассчитанные по методике [13] для метода введения идентификаторов до модернизации (контрольный вариант) и после нее. Критичным для идентификации является уровень $ВИ=0,01$. PeopleFullName – чувствительный атрибут ($ВИ=1$ исключен при модернизации). HospitalID – нечувствительный атрибут ($ВИ=0,0027$, необходимый идентификатор). PeopleBirthday – нечувствительный атрибут ($ВИ=0,0014$ исключен при модернизации, так как пара атрибутов HospitalID+PeopleBirthday является чувствительной: $ВИ=0,75$).

Рис. 2. Изменение вероятности идентификации при внедрении метода идентификаторов

Показатель степени обезличивания является интегральной характеристикой для всей совокупности атрибутов физических лиц. До модернизации его значение равно нулю из-за наличия чувствительных атрибутов в базе данных. После модернизации значение степени обезличивания достигло 0,997, что явно лучше критического уровня 0,99.

Таким образом, применение обезличивания (с обязательной возможностью деобезличивания) может обеспечить эффективную защиту персональных данных. Опыт внедрения методов обезличивания в России за последние 5 лет нельзя назвать достаточным, но он позволяет оценить некоторые тенденции. Наиболее важной представляется зависимость эффективности обезличивания от метода обезличива-

ния. Несмотря на отсутствие методической базы оценки эффективности, достаточно уверенно можно утверждать, что наиболее эффективным с точки зрения обеспечения безопасности ПД является метод введения идентификаторов.

Надежные оценки экономической эффективности внедрений делать сложно, поскольку цель получения экономического эффекта преследовалась не во всех случаях. С целью более широкого внедрения различных методов обезличивания персональных данных предполагается совершенствование методики оценки эффективности обезличивания в направлении, позволяющем сформировать соответствующую нормативную базу.

Работа выполнена при поддержке Правительства РФ (Постановление № 211 от 16.03.2013 г.), соглашение № 02.А03.21.0011.

Литература

1. Федеральный закон от 27.07.2006 «О персональных данных» № 152-ФЗ [Электронный ресурс]. – Режим доступа: http://www.consultant.ru/document/cons_doc_LAW_61801/, свободный (дата обращения: 14.05.2018).
2. Приказ Федеральной службы по надзору в сфере связи, информационных технологий и массовых коммуникаций от 05.09.2013 № 996 «Об утверждении требований и методов по обезличиванию персональных данных» [Электронный ресурс]. – Режим доступа: <http://54.rkn.gov.ru/protection/acts/p13580/>, свободный (дата обращения: 14.05.2018).
3. Вишнякова О.А. Формат обмена данными в системе сбора и обработки биометрических образцов / О.А. Вишнякова, Д.Н. Лавров // Информационные ресурсы в образовании: матер. междунар. науч.-практ. конф. – Нижневартовск: Изд-во Нижневарт. гос. ун-та, 2013. – С. 146–149.
4. Волокитина Е.С. Метод и алгоритмы гарантированного обезличивания и реидентификации субъекта персональных данных в автоматизированных информационных системах: дис. ... канд. техн. наук. – СПб.: Изд-во Санкт-Петербургского нац. исслед. ун-та информационных технологий, механики и оптики, 2013. – 183 с.
5. Пат. RU 103 414 U8, МПК G06F 17/40 (2006.01). Система взаимодействия разделенных баз персональных данных информационной системы / Д.Н. Иванов (RU), Е.Ю. Мищенко (RU). – № 2010149391/08; заявл. 02.12.2010; опубл. 10.04.2011. Бюл. № 10. – 2 с.
6. Кошкарров А.А. Система управления базами данных льготного лекарственного обеспечения в Краснодарском крае с использованием облачных технологий / А.А. Халафян, А.А. Кошкарров // Политематический сетевой электронный научный журнал Кубан. гос. аграрного ун-та. – Краснодар: Изд-во КубГАУ, 2015. – № 109 (05). – С. 451–467.
7. Кучин И.Ю. Обработка баз данных с персонифицированной информацией для задач обезличивания и поиска закономерностей: дис. ... канд. техн. наук. – Астрахань: Изд-во Астрах. гос. техн. ун-та, 2012. – 132 с.
8. Бондаренко К.О. Универсальный быстродействующий алгоритм процедур обезличивания данных / К.О. Бондаренко, В.А. Козлов // Изв. ЮФУ. Технические науки. – Ростов/н/Д: Изд-во ЮФУ, 2015. – № 11 (172). – С. 130–142.
9. Денисов М.И. Защита персональных данных в информационной системе медицинского учреждения методом обезличивания / М.И. Денисов, К.А. Чехонин // Научно-техническое и экономическое сотрудничество стран АТР в XXI веке. – Хабаровск: Изд-во Дальневосточного гос. ун-та путей сообщения, 2013. – Т. 1. – С. 229–232.
10. Воронин В.В. Защита персональных данных в информационных системах методом обезличивания / В.В. Воронин, Н.Л. Нехай // Информационные технологии XXI века: сб. науч. тр. – Хабаровск: Изд-во Тихоокеанского гос. ун-та, 2017. – С. 479–483.
11. Трифонова Ю.В. Возможности обезличивания персональных данных в системах, использующих реляционные базы данных / Ю.В. Трифонова, Р.Ф. Жаринов // Доклады ТУСУР. – 2014. – № 2 (32). – С. 188–194.
12. Ажмухамедов И.М. Системный подход к обеспечению конфиденциальности обезличенных персональных данных в учреждениях здравоохранения / И.М. Ажмухамедов, Р.Ю. Демина, И.В. Сафаров // Современные проблемы науки и образования. – 2015. – № 1-1 [Электронный ресурс]. – Режим доступа: <http://www.science-education.ru/ru/article/view?id=18610/>, свободный (дата обращения: 16.11.2018).
13. Мищенко Е.Ю. Количественный анализ процедуры обезличивания персональных данных. Метод введения идентификаторов / Е.Ю. Мищенко, А.Н. Соколов // Вестник ЮУрГУ. Сер.: Компьютерные технологии, управление, радиоэлектроника. – Челябинск: Изд. центр ЮУрГУ, 2015. – Т. 15, № 3. – С. 18–25.
14. Мищенко Е.Ю. Количественный анализ процедуры обезличивания персональных данных. Метод изменения состава или семантики / Е.Ю. Мищенко, А.Н. Соколов // Вестник УрФО. Безопасность в информационной сфере. – Челябинск: Изд. центр ЮУрГУ, 2016. – № 1 (19). – С. 30–38.
15. Мищенко Е.Ю. Количественный анализ процедуры обезличивания персональных данных. Метод перемешивания / Е.Ю. Мищенко, А.Н. Соколов // Вестник УрФО. Безопасность в информационной сфере. – Челябинск: Изд. центр ЮУрГУ, 2016. – № 3 (21). – С. 30–37.

Мищенко Евгений Юрьевич

Ст. преподаватель каф. защиты информации
Федерального государственного автономного образовательного учреждения высшего образования
«Южно-Уральский государственный университет»
(Национальный исследовательский университет))
ФГАОУ ВО «ЮУрГУ (НИУ)»
Ленина пр-т, 76, г. Челябинск, Россия, 454080
Тел.: +7 (351) 267-93-55
Эл. почта: mishchenkoei@susu.ru

Соколов Александр Николаевич

Канд. техн. наук, доцент, зав. каф. защиты информации
ФГАОУ ВО «ЮУрГУ (НИУ)»
Ленина пр-т, 76, г. Челябинск, Россия, 454080
Тел.: +7 (351) 267-93-55
Эл. почта: sokolovan@susu.ru

Mishchenko E.Yu., Sokolov A.N.

Implementation algorithms for the methods of personal data depersonalization in distributed information systems

The article deals with various methods of personal data protection, based on the use of depersonalization procedures, as well as methods for assessing the efficiency of depersonalization.

The relevance of the use of depersonalization is conditioned by the fact that unlike other methods of personal data protection (i.e. the encryption), the cost of depersonalization does not depend on the number of points of the processing of personal data. That can provide certain economic benefits. The authors describe the characteristics of different depersonalization methods that have an important impact on the implementation of those methods and also give an overview of the implemented algorithms in various fields of personal data processing. In addition, the authors present the results of evaluating the effectiveness of depersonalization of personal data in the health sector obtained by the method of identifiers. As a result of the implementation, a sufficient degree of protection of personal data was achieved, and a significant economic effect was obtained.

Keywords: probability of identification, information system, method of introducing identifiers, method of changing the composition or semantics, method of mixing, depersonalization of personal data, degree of depersonalization.

doi: 10.21293/1818-0442-2019-22-1-7-66-70

References

1. Federal'nyi zakon ot 27.07.2006 № 152 «O personal'nyh dannyh» (In Russ.). Available at: http://www.consultant.ru/document/cons_doc_LAW_61801/ (accessed: May 14, 2018).
2. Prikaz Federal'noj sluzhby po nadzoru v sfere svyazi, informacionnyh tekhnologij i massovyh kommunikacij ot 05.09.2013 № 996 «Ob utverzhdenii trebovanij i metodov po obezlichivaniyu personal'nyh dannyh» (in Russ.). Available at: <http://54.rkn.gov.ru/protection/acts/p13580/> (accessed: May 14, 2018).
3. Vishnyakova O.A., Lavrov D.N. The data exchange format in the system for the collection and processing of biometric samples. Information resources in education: materials of International scientific-practical conference. Nizhnevartovsk: Publishing house of Nizhnevartovsk State University, pp. 146–149, 2013 (in Russ.).
4. Volokitina E.S. Method and algorithms of guaranteed depersonalization and reidentification the subject of personal data in automated information systems. Thesis of candidate of technical Sciences. Saint-Petersburg: Publishing house of Saint-Petersburg National Research University of Information Technologies, Mechanics and Optics, 2013, 183 p. (in Russ.).
5. Pat. RU 103 414 U8, МПК G06F 17/40 (2006.01). Ivanov D.N. (RU), Mishchenko E.Yu. (RU). The communication system of the divided personal databases of the information system. № 2010149391/08, decl. 02 Dec. 2010, publ. 10 Apr. 2011, Bull. no 10, 2 p. (in Russ.).
6. Khalafyan A.A., Koshkarov A.A. The database management system of preferential drug supply in the Krasnodar region with the use of cloud technologies. *Polythematic network electronic scientific journal of the Kuban state agrarian University. Krasnodar: Publishing house of the Kuban State Agrarian University*, 2015, no. 109 (05), pp. 451–467. (in Russ.).
7. Kuchin I.Yu. Processing databases with personal information for the purposes of depersonalization and finding patterns, Thesis of Candidate of Engineering. Astrakhan: publishing house of the Astrakhan state technical University, 2012, 132 p. (in Russ.).
8. Bondarenko K.O., Kozlov V.A. A universal fast algorithm procedures for the depersonalization of data. *Izvestiya YuFU. Technical Sciences. Rostov-on-Don: publishing house of the Southern Fed. University press*, 2015, no. 11 (172), pp. 130–142 (in Russ.).
9. Denisov M.I., Chekhonin K.A. Protection of personal data in the information system of medical institution method of depersonalization. *Scientific, technical and economic cooperation of Asia Pacific region in the twenty-first century. Khabarovsk: Publishing house of far Eastern State University of Railways*, 2013, vol. 1, pp. 229–232 (in Russ.).
10. Voronin V.V., Nekhai N.L. Personal data protection in information systems by the method of depersonalization. *Information technologies of the XXI century: collection of scientific works. Khabarovsk: Publishing house of Pacific State University*, 2017, pp. 479–483 (in Russ.).
11. Trifonova Yu.V., Zharinov R.F. The possibility of depersonalization of personal data in systems that use relational databases. *Proceedings of TUSUR University. Tomsk: Publishing house Tomsk State University of Control Systems and Radioelectronics*, 2014, no. 2 (32), pp. 188–194 (in Russ.).
12. Azhmukhamedov I.M., Demina R.Yu., Safarov I.V. Systematic approach to ensuring the confidentiality of depersonalized personal data in health care institutions // Modern problems of science and education, 2015. – № 1-1 (In Russ.). Available at: <http://www.science-education.ru/ru/article/view?id=18610/> (accessed: November 16, 2018).
13. Mishchenko E.Yu., Sokolov A.N. Quantitative Analysis of the Depersonalization Procedure. Method of Identifiers. *Bulletin of the South Ural State University. Ser. Computer Technologies, Automatic Control, Radio Electronics*, 2015, vol. 15, no. 3, pp. 18–25 (in Russ.).
14. Mishchenko E.Yu., Sokolov A.N. Quantitative analysis of the depersonalization procedure. Method of composition or semantics modification. *UrFR Newsletter. Information Security*, 2016, no. 1 (19), pp. 30–38 (in Russ.).
15. Mishchenko E.Yu., Sokolov A.N. Quantitative analysis of the depersonalization procedure. Method of mixing. *UrFR Newsletter. Information Security*, 2016, no. 3 (21), pp. 30–37 (in Russ.).

Evgeniy Yu. Mishchenko

Senior Lecturer, Department of Information Security,
South Ural State University (National Research University)
76, Lenin pr., Chelyabinsk, Russia, 454080
Phone: +7 (351) 267-93-55
Email: mishchenkoei@susu.ru

Alexander N. Sokolov

Candidate of Engineering, Assistant Professor,
Chief of Department of Information Security,
South Ural State University (National Research University)
76, Lenin pr., Chelyabinsk, Russia, 454080
Phone: +7 (351) 267-93-55
Email: sokolovan@susu.ru