УДК 004.056.5 Вестник УрФО № 1(19) / 2016, стр. XX–XX

Мищенко Е. Ю., Соколов А. Н.

КОЛИЧЕСТВЕННЫЙ АНАЛИЗ ПРОЦЕДУРЫ ОБЕЗЛИЧИВАНИЯ ПЕРСОНАЛЬНЫХ ДАННЫХ. МЕТОД ИЗМЕНЕНИЯ СОСТАВА ИЛИ СЕМАНТИКИ

Обезличивание – способ обработки персональных данных, целью которого является приведение этих данных в защищенное состояние, которое не позволяет злоумышленнику использовать их во вред физическому лицу. Результат обезличивания персональных данных зависит от их содержания и применяемого метода обезличивания. Нормативные акты определяют несколько методов обезличивания, но все они описываются качественными критериями. В статье производится количественный анализ одного из методов обезличивания – метода изменения состава или семантики. Предлагается вариант технической реализации данного метода, включая решение проблемы необходимого и достаточного идентификационного набора изменяемых атрибутов, определение требований к правилам изменения, а также рассмотрение возможных способов реализации таких требований. На основе реального примера производится оценка эффективности метода по различным критериям. В том числе по техническим критериям (невозможность идентификации, с одной стороны, и возможность деобезличивания с применением заданных правил, с другой стороны), а также по экономическим критериям (окупаемость). На базе показателей вероятности идентификации и степени обезличивания персональных данных приводятся рекомендации по повышению эффективности данного метода обезличивания персональных данных.

Ключевые слова: *персональные данные, обезличивание персональных данных, метод изменения состава или семантики.*

Mishchenko E. Yu., Sokolov A. N.

QUANTITATIVE ANALYSIS OF THE DEPERSONALIZATION PROCEDURE. METHOD OF COMPOSITION OR SEMANTICS MODIFICATION

Depersonalization is the way of personal data processing for the purpose of transforming data to protected status, in order to prevent disturber use it to damage the person. The result of depersonalization is depending on content of the personal data and the depersonalization

method also. Standard acts define some methods of depersonalization, but all of them are describing by qualitative criterions. This article makes the quantitative analysis of one of the depersonalization methods – the method of composition or semantics modification (MOD Method). Proposed variant of technical realization of MOD Method solves the problem of the identification attribute set, which is necessary and sufficient for the modification, define the requirements for modification rules, and also describes the methods to realize them. On practical example the performance evaluation of MOD Method is made by using some criterions, including the technical criterions (identification impossibility on the one side and reconstruction possibility by means of modification rules on the other side), and commercial criterions (economic return). On the base of identification probability and depersonalization degree some recommendations of efficiency enhancement are proposed.

Keywords: *personal data, depersonalization, method of composition or semantics modification.*

Приказом Роскомнадзора[1] предусмотрено четыре метода обезличивания, реализующих различные качественные принципы изменения базы персональных данных:

1) Метод введения идентификаторов реализует принцип подстановки в БД абстрактного идентификатора вместо группы идентифицирующих атрибутов. Эта группа хранится в секрете, а остальные данные становятся обезличенными. В статье «Количественный анализ процедуры обезличивания персональных данных. Метод введения идентификаторов»[2] этот метод рассмотрен подробно с точки зрения количественных критериев.
2) Метод изменения состава или семантики реализует принцип подстановки в БД абстрактных значений (в том числе с иной структурой) вместо значений идентифицирующих атрибутов. При этом БД становится обезличенной, а в секрете хранится алгоритм восстановления прежних значений атрибутов. Необходимо отметить, что данный метод в общем случае допускает необратимую модификацию БД, вплоть до уничтожения идентифицирующих атрибутов. Но поскольку нас интересует возможность восстановления ПД, мы будем рассматривать данный метод только в таком аспекте. Определение значений количественных критериев для данного метода обезличивания является целью данной статьи.
3) Метод перемешивания реализует принцип подстановки в БД чужих значений (из другой записи этой же БД) вместо значений любых атрибутов. При этом БД становится обезличенной, а в секрете хранится алгоритм восстановления прежних значений атрибутов. Отличие от предыдущего метода состоит в том, что изменяется не само значение атрибута, а его место. Определение значений количественных критериев для данного метода обезличивания выходит за рамки данной статьи и будет рассмотрено в дальнейшем.
4) Метод декомпозиции реализует принцип разделения БД на произвольное количество групп атрибутов с дальнейшим раздельным хранением этих групп. Предполагается также создание некоторых таблиц связи, которые будут храниться в секрете. Данный метод можно рассматривать как некий симбиоз предыдущих методов: с одной стороны, нельзя обойтись без изменения порядка расположения внутри каждой группы атрибутов, так как БД в целом не будет обезличена и таблицы связей не имеют смысла, а изменение порядка приводит к методу перемешивания (но с раздельным хранением). С другой стороны, если разделить БД всего на две связанные части, то получим метод введения идентификаторов (с сомнительным дополнением в виде таблицы связей). Мы не утверждаем, что данный метод не имеет перспективы, просто его количественные критерии аналогичны методам перемешивания либо введения идентификаторов.

Каждый из указанных методов подразумевает неограниченный доступ к обезличенной части, с одной стороны, и ограниченный

(санкционированный) доступ к зависящей от метода защищаемой части (секрет метода), с другой стороны. То есть независимо от метода обезличивания ПД должны существовать рабочие места, являющиеся частью защищаемой ИСПДн, на которых хранится эта секретная часть, происходят собственно процессы обезличивания и деобезличивания. Возможны варианты, когда секретом являются не собственно ПД и даже не программное обеспечение для их модификации (и то, и другое занимает много места и должно храниться стационарно на месте обработки), а некий очень малый набор параметров или коэффициентов, ключевым образом влияющих на алгоритм обработки (хранится на внешнем носителе).

В предыдущей статье[2] показаны две возможные цели процесса обезличивания с точки зрения технологии:

• передача обезличенной информации по незащищенным каналам связи (от одной группы защищенных рабочих мест к другой такой же группе);

• обработка обезличенной информации на незащищенных рабочих местах (с двусторонней передачей на защищаемые рабочие места).

Также показано, что метод введения идентификаторов в целом более эффективен для достижения второй цели.

Ниже приводится исследование следующего метода обезличивания – метода изменения состава или семантики.

1. Описание метода. В соответствии с Приказом Роскомнадзора[1] метод изменения состава или семантики реализуется путем обобщения, изменения или удаления части сведений, позволяющих идентифицировать субъекта, и создания некоторого правила модификации этих сведений. Вариант удаления придется исключить из-за невозможности деобезличивания, поэтому учитываем только различные способы модификации значений атрибутов. То есть после применения данного метода единая база (БД), в отличие от метода введения идентификаторов, не распадается на части, но возникают следующие особенности:

1) БД преобразуется в обезличенную базу с тем же объемом, но не идентичной структурой, в которой некий набор идентифицирующих физическое лицо (ФЛ) атрибутов становится набором неких более или менее абстрактных значений. Менее абстрактными можно считать изменения типа перевода значения на иностранный язык или деление числа на 10, более абстрактным можно считать использование любого кодирования, в т.ч. шифрования. Остальные атрибуты, не значимые с точки зрения идентификации, но определяющие суть обработки, остаются неизменными.
2) Создаются алгоритмы прямого и обратного преобразования значений атрибутов. Алгоритмы должны быть формализованы для автоматизированного (ПО) либо ручного (текстовая инструкция) использования.

В данном процессе модификации БД необходимо решить три проблемы:

1) Какие атрибуты включить в группу модифицируемых.
2) Какими свойствами должно обладать модифицированное значение.
3) Какими свойствами должен обладать алгоритм модификации.

1.1. Атрибуты модифицируемой группы. Набор атрибутов, подлежащих модификации, аналогично методу введения идентификаторов должен соответствовать двум требованиям:

• быть достаточным для идентификации конкретного ФЛ (интегральный показатель ВИ для данного набора должен быть равен 1).

• быть необходимым для надежного обезличивания прочих данных, не включенных в эту группу (интегральный показатель ВИ для любого набора из оставшихся реквизитов должен быть не просто меньше 1, а нормативно меньше).

В статье «Количественные критерии идентификации физического лица при обезличивании персональных данных»[3] обосновано, что для идентификации ФЛ в объеме 1 млн записей достаточным является набор «фамилия» + «дата рождения», но в то же время набор таких атрибутов как «имя», «адрес проживания», «номер телефона», «место работы», может дать $ВИ=1$ сам по себе - без атрибута «фамилия».

Следовательно, в модифицируемую группу должны быть включены все атрибуты, по которым возможно идентифицировать ФЛ с ВИ больше $ВИ_{норм}$. В частности, в группу модифицируемых должны войти все так называемые официальные реквизиты (ИНН, СНИЛС и т.д.)

Однако, необходимо учитывать, что реальная обработка ПД автоматизированным способом производится в рамках нескольких таблиц базы данных реляционного типа, одна часть которых является справочниками (условно постоянные), а другая часть – изменяемые данные функционального характера (переменные). Эти части связаны посредством специальных служебных идентификаторов. Очевидно, что эти связующие идентификаторы не могут быть модифицированы независимо в различных таблицах, и таким образом не должны включаться в модифицируемую группу. Следовательно, если они являются идентифицирующими (имеют реальный смысл), то должны быть переведены в абстрактный вид до начала применения метода.

1.2. Требования к модифицированным значениям. Из определения метода следует, что целью модификации идентифицирующего атрибута является превращение его в непригодный для идентификации вид. Поскольку модификации подвергается не один атрибут, а группа, вышеприведенное требование можно сформулировать более строго: интегральная вероятность идентификации ФЛ для группы модифицированных реквизитов в целом должна быть меньше нормативного значения. При этом модифицированные значения могут сохранять семантику (например, значение «Москва» атрибута «город» модифицировалось в значение «М----а»), а могут стать абстрактным набором символов. Но какие изменения являются минимально необходимыми? Очевидно, что ВИ для атрибута достаточно абстрактного вида практически равна нулю, но чем более модифицированными (абстрактными) будут измененные атрибуты, тем больше усилий потребуется в дальнейшем для их восстановления. Поэтому вполне достаточными могут оказаться даже незначительные изменения.

1.3. Требования к алгоритму модификации. Из определения метода следует, что алгоритм может изменять такие характеристики БД, как состав и семантику. С точки зрения технологии обработки в рамках реляционной базы данных модификация – это копирование значений полей (или их частей, в т.ч. преобразованных любым способом) из исходной БД в какое-либо поле конечной БД. Любое поле БД содержит атрибут и имеет три характеристики (которые могут отличаться в конечной БД в сравнении с исходной):

• структурную (имя поля в БД, как идентификатор единицы обработки), в статье3 обозначенную как собственно атрибут АБ;

• семантическую (тип данных поля, как показатель формата хранения), в статье3 обозначенную как название атрибута НБ;

• содержательную (значение поля), в статье3 обозначенную как значение атрибута ЗБ.

При рассмотрении вариантов изменений (реализаций алгоритмов) необходимо учитывать, что нас интересует обратимость обезличивания, то есть для предлагаемого алгоритма модификации должен быть предусмотрен и обратный алгоритм восстановления значений атрибутов.

Под изменением состава надо понимать следующие варианты изменения структуры:

• добавление полей. Прямая цель – внесение в конечную БД дополнительных ложных значений атрибута (дезинформация), вспомогательная – для копирования исходного значения (или его части) из одного поля в несколько конечных (выглядит как разделение одного атрибута на несколько частей);

• удаление полей. Прямая цель исключается, как необратимая операция (потеря атрибута), вспомогательная – необходима при слиянии нескольких исходных атрибутов в один конечный;

• перемещение полей. Условная операция - расположение (перечисление в структуре) аналогичных полей в конечной БД в порядке, отличном от исходной БД, создающая эффект перемещения. Прямая цель сама по себе не имеет смысла, вспомогательная – необходима при разделении атрибутов;

• разделение полей. Имеет смысл при совместном использовании с добавлением и перемещением атрибутов в конечной БД;

• слияние полей. Имеет смысл при первоначальном разделении и перемещении полей в конечной БД.

Наиболее сложным вариантом изменения состава является преобразование исходной БД в конечную, в которой количество полей идентифицирующих атрибутов равно суммарному количеству их символов в исходной БД (размер соответствующих полей конечной БД – один символ), причем эти поля расположены в произвольном порядке.

Под изменением семантики надо понимать изменение содержимого атрибута. Оно принципиально отличается от всех предыдущих изменений тем, что перед копированием из исходного поля в конечное преобразуется

само значение атрибута. Преобразование может быть любым.

Если алгоритм модифицирует только структуру (состав) БД, то секретом алгоритма будет правило (таблица) преобразования исходных полей в конечные, то есть соответствия от простого вида «имя - имя» до наиболее сложного «имя - группа имен+группа размеров».

Алгоритм, модифицирующий значения атрибутов, – самый сложный, возможно, его секретом будет лишь небольшой набор параметров формулы преобразования (например, ключ шифрования). Конечно, его можно еще более усложнить, применяя в совокупности с модификацией структуры исходной БД.

2. Оценка эффективности метода. В Приказе Роскомнадзора[1] определены два термина, которые могут быть приняты критериями эффективности. Это «анонимность», как критерий функциональности (в статье[3] для него введен количественный аналог - степень обезличивания СО, которая связана с ВИ формулой $СО = 1 - ВИ_{макс}$), и «применимость», как критерий технической реализуемости, отражающий возможность обработки обезличенной базы без предварительного деобезличивания, т.е. защищенной обработки в с применением «дополнительной информации» и незащищенной обработки - без ее применения. В рамках метода изменения состава или семантики такой «дополнительной информацией» является секретный алгоритм (либо его параметры в виде таблиц или ключей преобразований).

2.1. Эффективность анонимности. В предыдущей статье[2] мы производили оценку значения СО на конкретном примере: злоумышленник ищет ФЛ (фамилию, имя, отчество, адрес и пр.) на основании известной ему информации о его автомобиле (внешнем виде) при условии свободного доступа к обезличенной БД регистрации всех автомобилей нашей страны. Только теперь мы будем считать, что эта БД обезличена методом изменения состава или семантики (рассмотрим оба этих варианта), и там есть в открытом виде вся информация об автомобилях, но все идентифицирующие атрибуты владельцев изменены в рамках метода).

Кроме доступа к БД злоумышленник может использовать следующую информацию, которую можно получить из открытых источников:

1) Злоумышленнику известен регион, в котором ФЛ эксплуатирует свой автомобиль (средний регион нашей страны - с населением 2 млн. человек, областной центр - с населением 1 млн. человек);
2) Возраст ФЛ – от 18 до 60 лет;
3) Количество ДТП в год по региону – 3 тыс. при количестве автомобилей – 800 тыс., по областному центру – 2 тыс. при количестве автомобилей – 500 тыс.
4) Для оценки максимального значения ВИмакс, примем критерий «актуальность идентификации» (т.е. время злоумышленнику на поиск) равным 30 дням (имея в виду, что реальное его значение не более 3 дней).

Из первых двух условий следует, что по возрасту водителями в данном регионе могут быть 1 млн. человек, а в областном центре – 500 тыс. ($ВИ$=1/1000000 и $ВИ_{макс}$=1/500000 – облегчим задачу злоумышленнику).

Третье и четвертое условия позволяют оценить вероятность того, что искомый автомобиль можно будет из-за ДТП обнаружить в ограниченном количестве известных мест (пункты регистрации ДТП, страховые компании, автосалоны по ремонту). Если принять, что в ДТП участвуют 2 автомобиля, то для региона вероятность попадания в ДТП конкретного автомобиля равна (3/800=1/266)*2=1/133 за год, а за 30 дней («актуальность идентификации») – 1/133/12=1/1596. Для областного центра эта вероятность будет 1/1500. Но для определения $ВИ$ надо учесть $К$ = «ограниченное количество известных мест». Если в областном центре 5 пунктов регистрации ДТП, то $ВИ_{макс}$=1/1500/5=1/7500. Это ВИ без использования атрибутов из БД. Посмотрим, чем поможет злоумышленнику информация из обезличенной базы.

В состав маркера поиска ($МП$) войдут атрибуты: $НМ1$ = «марка», $НМ2$ = «модель», $НМ3$ = «цвет кузова», $НМ4$ = «государственный номер» - их можно надежно определить по внешнему виду. В обезличенной базе регистрации автомобилей в открытом виде есть все эти атрибуты и еще многие другие: «дата регистрации», «место регистрации», «наименование автосалона-продавца», реквизиты договора продажи, свидетельства о регистрации, полиса ОСАГО и пр. При обезличивании по методу введения идентификаторов все идентифицирующие атрибуты просто от-

сутствовали, и при этом $ВИ_{макс}$ было равно 1/1500 ($СО$=1-1/1500=0,9993), а в рассматриваемом методе все эти атрибуты есть, из-за чего $ВИ_{макс}$ может только повыситься. Величину $ВИ$ будет определять возможность идентификации ФЛ по модифицированным атрибутам. Выше мы уже рассмотрели набор модифицируемых атрибутов, но в качестве наиболее приоритетных для идентификации злоумышленник, очевидно, выберет фамилию, имя и адрес ФЛ, а уже потом – прочие атрибуты.

Рассмотрим первый вариант изменений – изменение состава. Как показано выше, наиболее перспективный способ – разделение полей. Если принять среднюю длину слова в русском языке 7 символов, можно рассчитать, что разделение атрибута «фамилия» на части по одному символу с произвольным их размещением в структуре обезличенной БД, то теоретическое количество сочетаний этих символов будет равно 5000. $ВИ$=1/5000 – отличный результат, но фактический эксперимент с фамилией автора этой статьи дал всего около 20 возможных фамилий. Однако, если дополнительно подобной процедуре подвергнуть «имя» (не говоря уже об «отчестве») ФЛ и перемешать его с символами фамилии, грубый расчет покажет $ВИ$ менее 1/40000, но это достижение будет практически бесполезным, если не модифицировать надежно атрибут «адрес» (название улицы, номер дома и квартиры), так как знание улицы дает среднее значение $ВИ$=1/1000 (в городе-миллионнике – около 1 тысячи улиц, хотя такое же значение получается для одного многоквартирного дома), но наличие номера дома и квартиры значительно повышает $ВИ$ (например, из 5 цифр наберется не более 100 вариантов трехзначных квартир в двузначных домах). Одно понятно – разделение на символы атрибута адрес будет эффективным только вразброс с символами фамилии, имени, отчества. Вывод – изменение состава можно использовать для обезличивания только в самом сложном варианте посимвольного разделения атрибутов, любые упрощения алгоритма недопустимы.

Рассмотрим второй вариант изменений – изменение семантики. Существует три возможных способа изменений содержимого:

• добавление символов. Цель – дезинформация, но добавление букв в атрибут «фамилия», а тем более «имя» и «улица», далеко не всегда сохраняет смысл этих атрибутов. Добавление цифровых символов смысл атрибута сохраняет, но не может понизить ВИ меньше 1/1000 (известна улица), даже если символов добавить много;

• удаление символов. Опасная операция с точки зрения дальнейшего восстановления, а с точки зрения обезличивания – совсем не очевидная. Удаление первой буквы «фамилии» будет необратимым, а последней буквы – абсолютно не эффективным. Однако, возможен вариант применения данного способа при его совместном использовании с бумажным носителем. Предположим, что автомобиль остановил сотрудник ГИБДД, а у водителя нет документов на автомобиль. В реальной обстановке, чтобы подтвердить право на управление автомобилем, водитель может предъявить паспорт, но в нашем примере у сотрудника ГИБДД есть только обезличенная база, поэтому реквизиты из паспорта не помогут. Попробуем сократить «фамилию», «имя», «отчество» собственника до первых букв и посмотрим, чему равно $ВИ$. В списке 500 самых распространенных русских фамилий 34 фамилии начинаются на букву «М», то есть $ВИ$ равно 0,07, для татарских фамилий – $ВИ$=0,07, для украинских – $ВИ$=0,08. Независимо от национальности ФЛ в словаре «Российские фамилии в алфавитном порядке» (2014 г) для объема 250000 фамилий расчет показывает $ВИ$=0,08. Это значение показывает вероятность того, что случайный человек (независимо от пола) имеет фамилию на букву «М» (ресурс http://russkg.ru/index.php?catid=84:2012-12-02-23-13-33&id=4390:-500&Itemid=63&option=com_content&view=article). Аналогичные расчеты показывают, что вероятность имени на букву «Е» равна 0,023 (ресурсы http://imyarebenku.ru/man_names и http://imyarebenku.ru/woman_names), а вероятность мужского имени на букву «Ю» равна 0,008 (примем ее за вероятность отчества). Вероятность совпадения всех трех атрибутов менее 1/67000. Для более часто встречающихся инициалов вероятность гораздо выше (для инициалов «САН» - 1/5700!), но надо учитывать, что в поставленной задаче злоумышленник не знает инициалов и, хотя $ВИ$ увеличится вдвое из-за возрастных пределов, все равно ему удобнее для идентификации использовать открытые атрибуты. Что касается сотрудника ГИБДД, который знает инициалы из паспорта, то водителя с инициалами «МЕЮ» можно отпустить, а с инициалами «САН» придется задер-

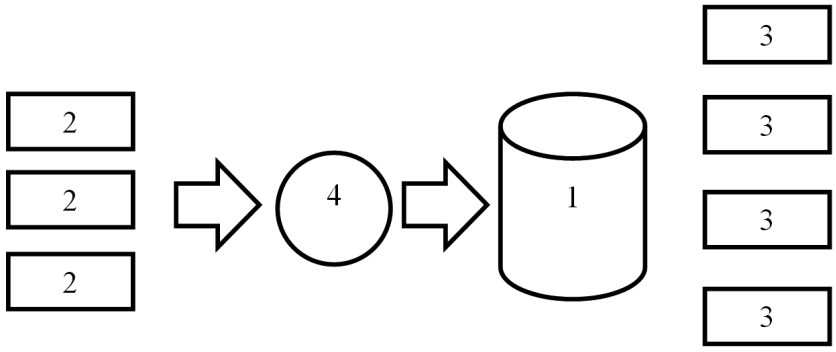

Рис. 1

жать. В целом данный способ изменения содержимого к сожалению не годится, так как позволяет злоумышленнику достаточно надежно определить количество автомобилей в собственности известного ему ФЛ;

• замена символов «один-в-один» (без изменения размера атрибута). Решающим образом зависит от алгоритма замены. Обычная перестановка символов атрибута (без их изменения) в произвольном порядке приводит нас к уже рассмотренному варианту разделения полей с их перестановкой, но ограниченному размерами самого атрибута, что является недостаточным. Любая кодировка, рассчитанная на секретный алгоритм (программный код в виде исполняемого модуля, хранящегося на рабочем месте оператора), например, в виде квадратного уравнения О=к*И2+л*И+м, где И - исходное значение атрибута, О – обезличенное значение, к, л, м – константы, легко вычисляется путем использования известных значений (злоумышленник найдет в обезличенной базе информацию о себе по номеру своего автомобиля и вычислит константы уравнения, т.е. ВИ данного способа =1). Любая кодировка, рассчитанная на секретные константы (таблица замен и т.п., хранящиеся на внешнем носителе) по мере ее усложнения приведет нас к шифрованию, то есть к использованию секретных ключей. Шифрование является абсолютно надежным способом, но даже при его использовании не удастся уменьшить ВИ по сравнению с использованием метода введения идентификаторов, конечно, если не применить полное шифрование всех атрибутов (в этом случае ВИ=0).

2.2. Эффективность применимости. Решающим критерием применимости рассматриваемого метода является техническая возможность его реализации. А уже при наличии технической возможности определяющую роль играют стоимость и сроки реализации.

На рис.1 приведена схема разделения обработки ПД на защищаемых и не защищаемых рабочих местах, где в качестве связующего звена используется межсетевой экран.

Цифрами на рис.1 обозначены:

1 – Обезличенные данные (сервер обезличенной базы, в свободном доступе, не защищается);

2 – Рабочее место оператора ПД с установленным алгоритмом модификации (входит в состав ИСПДн, защищается);

3 – Рабочее место оператора обезличенной базы (в свободном доступе, не защищается);

4 – Межсетевой экран, обеспечивает односторонний доступ к информации, направление которого указано стрелками.

В нашем случае с БД регистрации автомобилей все чувствительные к идентификации ПД владельцев (фамилия, имя, отчество, дата рождения, место рождения, адрес проживания, телефон, номер паспорта) будут модифицироваться и восстанавливаться только пользователями рабочего места 2 (регистратура). А в базе 1 будут храниться и обрабатываться модифицированные идентификаторы ФЛ и открытые данные о постановке и снятии с учета всех автомобилей любого ФЛ, и доступ к ним будет разрешен не только легальным пользователям рабочих мест 3, но и любым пользователям, которые смогут (в принципе без ограничений) получить физический доступ к базе 1 (либо с рабочего места 3 в отсутствие легального пользователя, либо подключив к базе 1 новое рабочее место).

Анонимность обеспечивается межсетевым экраном, который запрещает доступ к информации (ПД) и алгоритму модификации на рабочем месте 2 от рабочих мест на стороне обезличенной базы, то есть злоумышленник (как и любой пользователь рабочего места 3) не имеет доступа к алгоритму.

Для определения эффекта применимости показанной на рис.1 системы в целом необходимо оценить применимость ее обеих частей – защищаемой (состоит из необходимого количества рабочих мест 2) и не защищаемой (состоит из сервера 1 и большого количества рабочих мест 3).

Пользователь рабочего места 2 может решать как частную, так и общую задачу идентификации, и другие функции:

1) указав значение любого идентификатора ФЛ, модифицировать его и выдать запрос через экран 4, получить все данные конкретного ФЛ из обезличенной базы (сервер 1) и сопоставить их с этим ФЛ;
2) указав значения открытых атрибутов из обезличенной базы 1, получить список модифицированных идентификаторов ФЛ с заданными значениями атрибутов и восстановить их в виде ПД;
3) кроме того, только этот пользователь может добавлять записи о новых ФЛ и модифицировать для них идентификаторы, а также удалять ФЛ из БД полностью (т.е. всю информацию о ФЛ).

Таким образом применимость (техническая реализация функций) с точки зрения защищаемой системы не вызывает сомнений.

Пользователь рабочего места 3 может выполнять аналогичную рабочему месту 2 функцию 2), но не может сопоставить эту информацию конкретным ФЛ, функцию 1) он может выполнить, только имея уже модифицированный идентификатор, а функцию 3) он выполнять не может в принципе.

Что касается функции ввода существенной информации (ради которой и создана вся система обработки), она может выполняться только в рамках функции 2), т.е. пользователем 2, а пользователем 3 только при наличии модифицированного идентификатора. Очевидно, с точки зрения применимости рабочего места 3, складывается ситуация полностью аналогичная показанной в статье[2], где описывалось использование метода введения идентификаторов.

При первом посещении (постановка автомобиля на учет) владелец должен прийти к рабочему месту 2, представить свои ПД (показать паспорт) и получить идентификатор. По методу введения идентификаторов ФЛ получал идентификатор из таблицы соответствий, но и в нашем случае ему не нужно предоставлять весь набор модифицированных идентификаторов – достаточно дать абстрактный технологический идентификатор, который все равно будет нужен для связи таблиц БД. После получения идентификатора ФЛ-владелец идет к рабочему месту 3, представляет идентификатор и все данные об автомобиле. При повторном посещении (снятие с учета или постановка на учет другого автомобиля) владелец идет с идентификатором сразу к рабочему месту 3. Тип внешнего носителя, на котором хранится идентификатор ФЛ, и проблемы, связанные с его применением, также обсуждались в статье[2].

С первого взгляда оба метода обезличивания имеют одинаковую эффективность применения, но при расширении границ применения сразу будут видны различия:

1) Количество мест стационарного хранения «секрета метода». В методе введения идентификаторов – 1 (таблица соответствий), а в методе изменения состава и семантики – количество рабочих мест с установленным алгоритмом и установочных носителей (если это внешние носители с ключами алгоритма – то же самое количество, хотя украсть их сложнее), кроме того – алгоритм известен разработчику;
2) Частичная утечка «секрета метода». Даже при утечке части таблицы соответствий злоумышленник не получит доступа к остальной части и к ПД всех новых ФЛ, а в методе изменения состава и семантики – любая утечка – полная и на все время до смены алгоритма;
3) Нарушение доступности «секрета метода». При распределенной системе обработки географическая удаленность таблицы соответствий от обезличенной базы порождает сложности в технологии применения метода введения идентификаторов, а в методе изменения состава и семантики – БД всегда в целом виде;
4) Передача по каналам связи таблицы соответствий в открытом виде – невозможна, в отличие от модифицированной базы в методе изменения состава и семантики.

2.3. Экономическая эффективность. Аналогично методу введения идентификаторов, в нашем случае наибольшие финансовые затраты также потребуются на модернизацию программного обеспечения (ПО)

ИСПДн. И это будет серьезным препятствием, когда производителем ПО является не оператор ПД, а некая сторонняя организация.

В статье2 мы уже определяли стоимость создания системы защиты ИСПДн, состоящей из такого же количества рабочих мест, как в обезличенной базе. Без учета средств межсетевого экранирования и обнаружения вторжений, а также разработки эксплуатационных документов (для защищенной ИСПДн и для обезличенной базы эти затраты будут одинаковыми) для 100 рабочих мест мы получили сумму 700000 руб.

Если сравнить эти затраты со стоимостью модернизации структуры БД и ПО (не зависит от количества рабочих мест и составляет от 0 (ПО собственного производства) до 300000 руб.), то целесообразность обезличивания будет однозначной. Но при сравнении экономики двух методов обезличивания необходимо учитывать, что для метода изменения состава и семантики каждого рабочего места 2 требуется наличие внешнего носителя (для хранения «секрета метода»), стоимость которого составляет около 1000 руб.

Таким образом, рассмотренный метод изменения состава и семантики характеризуется большей сложностью технической реализации, чем метод введения идентификаторов. В то же время эффективность обезличивания критически зависит от алгоритма модификации и всегда теоретически ниже, чем у метода введения идентификаторов. Тем не менее, эффективность метода является вполне достаточной, а его преимуществом является большая гибкость.

Примечания

1. Приказ Роскомнадзора от 5.09.2013 г. № 996 «Об утверждении требований и методов по обезличиванию персональных данных» [электронный ресурс]. URL: http://www.garant.ru

2. Мищенко Е.Ю., Соколов А.Н. Количественный анализ процедуры обезличивания персональных данных. Метод введения идентификаторов // Вестник ЮУрГУ. Серия «Компьютерные технологии, управление, радиоэлектроника». 2015. Т. 15, № 3. С. 18–25.

3. Мищенко Е.Ю., Соколов А.Н. Количественные критерии идентификации физического лица при обезличивании персональных данных // Вестник УрФО. Безопасность в информационной сфере. — Челябинск: Изд. центр ЮУрГУ, 2014. № 1(11) С. 27–33.

References

1. The Federal service for supervision in the sphere of Telecom, information technologies and mass communications decree) No.996, 2013. About approving the demands and methods of depersonalization of personal data. URL: http://www.garant.ru

2. E.Yu. Mishchenko, A.N. Sokolov. 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.

3. E.Yu. Mishchenko, A.N. Sokolov. Quantitative criterions of the person Identification for Depersonalization of personal data. Bulletin of the Ural Federal district. Safety in the information sphere. Chelyabinsk, South Ural St. Univ. Publ., 2014, no.1(11), pp. 27-33.

Мищенко Евгений Юрьевич, старший преподаватель кафедры безопасности информационных систем, Южно-Уральский государственный университет, г. Челябинск. E-mail: Eug6303@mail.ru.

Соколов Александр Николаевич, канд. техн. наук, доцент, заведующий кафедрой безопасности информационных систем, Южно-Уральский государственный университет, г. Челябинск. E-mail: ANSokolov@inbox.ru.

E.Yu. Mishchenko, South Ural State University, Chelyabinsk, Russian Federation. E-mail: Eug6303@mail.ru,

A.N. Sokolov, South Ural State University, Chelyabinsk, Russian Federation. E-mail: ANSokolov@inbox.ru