ОПРЕДЕЛЕНИЕ ЭФФЕКТИВНОСТИ ОБЕЗЛИЧИВАНИЯ ПЕРСОНАЛЬНЫХ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ МОДЕЛИ НАРУШИТЕЛЯ

В статье рассматривается модель нарушителя, как средство проверки эффективности некоторых методов обезличивания персональных данных (ПД), основанных на искажениях состава и структуры базы ПД. Отсутствие методик определения эффективности обезличивания ПД является препятствием для широкого применения на практике процедуры обезличивания, как способа защиты ПД. Приведено описание модели нарушителя, а также алгоритм воздействия на обезличенную базу данных, целью которого является раскрытие алгоритма искажения состава (или структуры). Критерием успеха алгоритма воздействия является минимум итераций. В то же время, отсутствие успеха алгоритма модели нарушителя указывает на достаточную эффективность алгоритма метода обезличивания. В результате эксперимента получены рекомендации по использованию конкретных параметров для методов обезличивания.

Ключевые слова: модель нарушителя, метод изменения состава или семантики, метод перемешивания, обезличивание персональных данных, вероятность идентификации.

DETERMINATION OF THE EFFECTIVENESS OF ANONYMIZATION OF PERSONAL DATA USING THE INTRUDER'S MODEL

The article discusses the violator's model as a means of testing the effectiveness of some methods of anonymization of personal data (PD) based on distortions of the composition and structure of the PD database. The lack of methods for determining the effectiveness of anonymization of PD is an obstacle to the widespread use of the procedure of anonymization in practice as a way to protect PD. A description of the intruder's model is given, as well as an algorithm for influencing an impersonal database, the purpose of which is to disclose an algorithm for distorting the composition (or structure). The criterion for the success of the impact algorithm is the minimum of iterations. At the same time, the lack of success of the intruder model algorithm indicates the sufficient efficiency of the depersonalization method algorithm. As a result of the experiment, recommendations were obtained on the use of specific parameters for methods of depersonalization.

Keywords: intruder's model, method of changing the composition or semantics, mixing method, depersonalization of personal data, identification probability.

1. Введение

Обезличивание ПД имеет главной целью нейтрализацию попыток злоумышленника использовать эти ПД во вред физическому лицу. Главная особенность Российского законодательства о ПД в отличие от зарубежного 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. Эта возможность позволяет обрабатывать ПД после их восстановления из обезличенной формы.

В 2013 году Приказом Роскомнадзора [2] были определены четыре основных метода обезличивания: введение идентификаторов, изменение состава или семантики, декомпозиция и перемешивание.

Принципиальное отличие методов состоит в отношении к идентифицирующей информации:

– методы идентификаторов и декомпозиции основаны на отделении идентифицирующей информации от обезличенных данных, при этом идентифицирующая часть недоступна для злоумышленника во время хранения, но может быть доступна во время прочих сеансов обработки;

– методы изменения состава/семантики и перемешивания основаны на скрытии (искажении) расположения идентифицирующей информации в базе обезличенных данных, поэтому для них проблемой является уязвимость алгоритма во время сеанса работы. При этом идентифицирующая часть продолжает находиться в искаженной базе, поэтому может быть извлечена злоумышленником (например, путем перебора или вычисления вариантов);

Общей проблемой всех методов, важной с точки зрения затрат (финансовых), является необходимость модификации структуры базы данных и прикладного программного обеспечения до начала применения метода обезличивания, а для искажающих методов

для работы в прозрачном режиме также велика зависимость вычислительных затрат от сложности алгоритма. Особенно важным это становится в режимах модификации обезличенной базы.

Рассмотрим варианты реализации искажающих методов обезличивания персональных данных с учетом имеющегося опыта их внедрения и возможности их взлома с помощью модели нарушителя.

2. Искажающие методы обезличивания персональных данных и модель нарушителя

Анализ состояния проблемы является результатом исследования авторами имеющихся научных публикаций по теме обезличивания ПД в России, а также разработки модели нарушителя, позволяющей вычислить таблицу. Все варианты реализации искажающих методов можно условно разделить на две группы по применяемым методам обезличивания.

2.1. Алгоритмы реализации метода изменения состава или семантики.

Согласно Приказа [2] суть данного метода состоит в внесении обратимых изменений (искажений) в идентифицирующие атрибуты каждой записи ПД, без использования информации из других записей базы. При этом алгоритм искажений не регламентируется, это может быть замена символов на другие символы либо перемещение символов внутри строки. Изменения могут производиться по таблице или по формуле, при этом таблица необязательно должна быть одинаковой для разных записей базы.

Единственный обнаруженный вариант внедрения данного метода представлен в работе И.Ю. Кучина [3], в которой предлагается способ кодирования идентифицирующих атрибутов на базе разработанного алгоритма. Предлагаемый алгоритм кодирует каждую запись базы отдельным кодом с использованием значений идентификаторов и параметров операционной системы. Этот алгоритм был внедрен в сфере страхования, но вопрос обеспечения безопасности решается только при хранении персональных данных небольшого объема, только в хранилище на базе определенной операционной системы.

Однако, использование различных таблиц замен для разных записей базы большого объема представляется авторам не целесообразным с точки зрения производительности, поэтому в данной работе рассматрива-

ется возможность применения одной таблицы для всех записей.

2.2. Алгоритмы реализации метода перемешивания.

Согласно Приказа [2] суть данного метода состоит в обратимом изменении (искажении) положения идентифицирующих атрибутов каждой записи в другие записи, но с сохранением семантики перемещаемых атрибутов (их места в строке). При этом алгоритм искажений не регламентируется, группа записей, внутри которой производится перемещение, может быть произвольной по количеству. Изменения могут производиться по таблице или по формуле, при этом таблица необязательно должна быть одинаковой для разных групп записей базы.

Среди обнаруженных внедренных вариантов можно отметить работу В.В. Воронина и Н.Л. Нехай [4], в которой авторы разработали алгоритм с различными циклическими сдвигами для каждого атрибута-идентификатора и внедрили его в сфере обслуживания автотранспорта. В предложенном алгоритме перемешивание осуществляется не в группе записей (сегменте базы), а в группах разной величины для различных идентификаторов. Данный подход можно назвать несегментированным.

В работе И.П. Карповой [5] приводятся расчеты производительности несегментированного подхода, из которых можно сделать вывод о нецелесообразности его применения для баз большого объема.

В других разработанных алгоритмах, рассчитанных на базы ПД большого объема, производится предварительное разделение базы на равные сегменты, в границах которых и производится перемешивание. Например, в работе К.О. Бондаренко и В.А. Козлова [6], размер сегмента составляет 256 записей, в сегменте перемешиваются сначала полные строки, затем идентификаторы между строками. Параметры перемешивания задаются криптографическими методами.

Авторы данной работы являются сторонниками сегментированного подхода, но считают, что применение криптографических методов в значительной мере нивелирует экономическую эффективность любых методов обезличивания.

2.3. Модель нарушителя. В своей работе авторы использовали экспериментальную базу ПД объемом 310 тыс. физических лиц, где в качестве идентифицирующих атрибутов

использовались фамилия, имя, отчество, название улицы проживания, номер дома и номер квартиры, в сумме составляющие строку длиной 73 символа. В качестве алгоритмов искажения рассматривались следующие:

- перестановка символов внутри полной строки идентификаторов;
- перестановка бит внутри идентификатора;
- перемешивание полей (с сохранением структуры строки) внутри группы из 256 записей базы;
- перемешивание символов (с сохранением места в строке) внутри группы из 256 записей базы.
- формула (таблица) смещения принималась произвольной, но одинаковой для всех строк (для метода изменения состава) или сегментов записей (для метода перемешивания).

Модель нарушителя, разработанная для идентификации физических лиц в обезличенной базе, может быть названа «Моделью внедренного пользователя» и имеет следующие параметры:

- нарушитель имеет неограниченный доступ к обезличенной базе;
- нарушитель знает структуру (размер полей и записи в целом), но не знает точной семантики исходной базы;
- нарушитель знает, что искажению подверглись только идентификаторы, т.е. прочие данные не искажены;
- нарушитель не знает формулы (таблицы) обезличивания базы, за исключением того, что все данные одного лица содержатся в одной записи базы (для метода изменения состава) или в одном сегменте не более 256 записей (для метода перемешивания);
- нарушитель знает только ФИО искомого лица, хочет узнать другие идентификаторы и прочие данные этого лица;
- нарушитель знает, что в базе есть данные некоторого числа знакомых ему лиц, которые он знает, причем прочие (неискаженные) данные может найти в базе;
- количество лиц в обезличенной базе, полная информация о которых известна нарушителю - не менее 3 и не более 5;
- нарушитель не может разрабатывать программное обеспечение аналитического типа для вскрытия алгоритма искажений, но может пользоваться готовыми программными средствами (например, для поиска или сравнения символов).

Алгоритм действий нарушителя следующий:

- нарушитель находит (вручную или автоматизированным путем) в базе прочие (неискаженные) данные первого известного ему лица;
- нарушитель оценивает возможность ошибки (одинаковые прочие данные могут быть у нескольких лиц). Оценка производится по составу символов искаженной строки или составу полей в сегменте из 511 записей (по 255 в обе стороны от найденной);
- нарушитель составляет таблицу смещений символов в строке или полей в сегменте по известным ему символам/полям знакомого лица. Таблица составляется по первому найденному значению и неизбежно будет частично ошибочной из-за совпадающих символов/полей, но определенная ее часть будет абсолютно точной;
- нарушитель использует для устранения ошибочной части таблицы данные второго известного ему лица по описанному выше алгоритму. Для устранения оставшихся ошибок можно использовать третью запись и т.д.;
- если нарушитель не получил точную таблицу смещений при использовании всех знакомых ему записей, данная модель считается неэффективной;
- нарушитель использует полученную точную таблицу смещений для получения данных о любом лице в данной базе, в этом случае модель считается эффективной.

2.3.1. Перемешивание символов внутри строки.

При определении смещений символов в перемешанной строке длиной N путем по-символьного сравнения с значением известной строки проблему представляют совпадающие (повторяющиеся) символы. Анализ частотного распределения символов в идентификаторах должен показать, какие символы наиболее вероятно встречаются в общей строке идентификаторов не менее двух раз и каково их общее количество - m . По причине строго заданной длины идентификаторов более всего будет повторяющихся пробелов, но их условно можно считать незначимыми и игнорировать в расчетах. При использовании нарушителем второй записи необходимо учитывать следующее:

- вторая строка независима от первой, поэтому в ней также останется m ошибочных мест, но они не совпадут с ошибочными местами возникшими после первой записи;

- минимальное количество совпадений ошибок двух строк = 0 (эти случаи возможны, если удвоенное количество ошибок меньше длины строки (в нашем случае условие $m \cdot 2 < N$ соблюдается)), но полное несовпадение сразу заполняет таблицу смещений и нас не интересует;

- если одному ошибочному месту первой записи соответствует точное место второй записи, то количество ошибок уменьшается на одно место;

- хотя повторяющиеся символы отличаются друг от друга, даже если место одного повторяющегося символа в первой строке совпадает с местом другого повторяющегося символа во второй строке, сам факт ошибки останется, поэтому все повторяющиеся символы условно считаем одним ошибочным символом.

Для расчета наиболее вероятного количества совпадений была применена формула:

$$C_{\max} = M(F_v(c)), \quad \{1\}$$

где M - математическое ожидание, $F_v(c)$ - функция зависимости количества вариантов совпадений от количества совпадений c . Сама функция $F_v(c)$ является дискретной и представляет собой произведение количества различных вариантов расположения с ошибочных символов в строке удвоенной длины ошибок на количество различных вариантов расположения оставшихся ошибочных символов в оставшейся части удвоенной строки, т.е.:

$$F_v(c) = P_{(2m-1)c} * P_{(2m-1-c)(m-c)} \quad \{2\}$$

где m - количество ошибочных мест, $P_{ab} = a! / (b!(a-b)!)$ - количество различных вариантов расположения b символов в строке длиной a (см. [7]). Сама функция $F_v(c)$ является симметричной, математическое ожидание будет соответствовать середине диапазона значений $C_{\max} = C_{cp}$, хотя при большом диапазоне более правильным будет использовать полосу значений от $C_{\max} - D$ до $C_{\max} + D$, где D - среднеквадратичное отклонение.

Теоретически при использовании третьей записи количество ошибок уменьшается вдвое, но поскольку ошибочные символы все-таки разные, их повторяемость уменьшится пропорционально и может стать меньше двух. Т.е. эти символы перестанут быть повторяющимися, и реальное количество ошибок уменьшится в большей степени.

Кроме того, необходимо учитывать, что сочетаемость различных букв любого языка

не является равномерной. Анализ частотного распределения пар символов в идентификаторах должен показать, какие сочетания символов наиболее вероятно встречаются в общей строке идентификаторов. Этот факт в результате еще более уменьшает количество ошибок.

2.3.2. Перемешивание бит внутри идентификатора.

В работе Е.Ю. Мищенко [8], где приводятся результаты исследования влияния различных идентификаторов на процесс идентификации, показано, что наиболее вероятную длину атрибута «фамилия» можно принять равной 8 (см. табл.1), остальные - пробелы. Если для упрощения расчетов абстрагироваться от пробелов и от кодов страниц различных языков, то в фамилии содержится 64 бита. Наиболее вероятное распределение - 33 бита 1 и 31 бит 0. Применить тот же подход, что при перемешивании символов можно, но необходимо учесть следующее:

- все биты 1 и 0 являются значимыми, т.е. теоретически ошибочными могут быть все 64 места и предыдущий подход в прямом виде применять нельзя ($2^{64} > 64$);

- подход усложняется и применяется по частям - сначала учитываются совпадения битов 1 друг с другом, затем учитываются совпадения битов 0 друг с другом, затем совпадения 1 и 0, в итоге количества вариантов соответствующих сочетаний перемножаются, и рассчитывается полная функция $F_v(c)$ и ее математическое ожидание;

- поскольку есть только биты 1 и 0, то никакие дополнительные условия уменьшения количества совпадений не срабатывают.

2.3.3. Перемешивание полей внутри групп записей.

Поскольку сегментированный подход подразумевает перемешивание идентификаторов в группе из 256 записей, то применяя первую известную запись, нарушитель должен найти в базе один из идентификаторов, и приняв в качестве точки отсчета абстрактный номер записи (его наличие необходимо во всех алгоритмах перемешивания), отступить в обе стороны от него по 255 записей и вести поиск прочих идентификаторов в этой области из 511 записей. Успех поиска зависит от возможности повтора известных значений идентификаторов в этой области. В работе [8] рассчитаны наиболее вероятные количества повторений значений различных идентификаторов в уже рассмотренной эксперимен-

тальной базе (см. столбец Мат.ожидание в табл.1).

Таблица 1

Атрибут	Мат. ожидание	Вероятность идентификации
Фамилия	7,5	0,133
Имя	132	0,008
Отчество	183	0,005
Улица	26	0,038
№ дома	132	0,008
№ квартиры	59	0,017

Используя мат. ожидание М, полученное для полной базы (310000 записей), можно определить количество повторений в группе 511 записей $M_{повт}$ по формуле:

$$M_{повт} = (M-1) \cdot 510 / 310000,$$

а вероятность повторения того же идентификатора при использовании второй записи - это и есть вероятность идентификации по данному атрибуту (см. табл.1).

2.3.4. Перемешивание символов внутри группы записей.

Если группу из 511 записей по 73 символа представить в виде повернутой таблицы из 73 записей по 511 символов, то для определения смещения символа в такой строке можно попытаться применить подход, описанный в п.2.3.1. и рассчитать количество необходимых знакомых записей для нарушителя.

3. Результаты применения модели нарушителя

3.1. Перемешивание символов внутри строки.

Для идентификаторов экспериментальной базы авторами произведен анализ частотного распределения символов, результаты которого приведены на рис.1.

Рис. 1

В строке идентификаторов длиной N=73 символа наиболее вероятно повторятся

следующие символы: пробел - 34 раза, «а» - 5 раз, «в», «е», «и», «н», «о» - по 3 раза, «к», «л», «р» - по 2 раза. Не повторяются - 13 символов. Пробелы при сравнении игнорируются, следовательно, в формируемой таблице смещений останется m=26 ошибочных мест.

Рис. 2

Из рис. 2 видно, что Fv(c) имеет явный максимум при Стах=13, которое и является математическим ожиданием, а среднеквадратичное отклонение для нашего случая D = 2,5, т.е. наибольшая часть вариантов находится в диапазоне от с=11 до с=15. При таком значении с применение второй знакомой записи нарушителем было неизбежно. Поскольку ошибочные символы все-таки разные, их повторяемость уменьшается пропорционально и становится меньше двух (для символов «к», «л», «р»). Повторяющихся символов стало меньше - совпадения сократились в три раза. Авторы также провели анализ сочетаемости букв в экспериментальной базе, который показал, что более 40% сочетаний являются маловероятными и даже невозможными (см. рис. 3).

Рис. 3

В результате количество совпадений

уменьшилось до 2-3х. При таком значении с использование третьей знакомой записи нарушителем может оказаться необязательным. А ее использование приведет к решению однозначно.

3.2. Перемешивание бит внутри идентификатора.

Применение аналогичного подхода к расчетам сначала для битов 1, затем для битов 0, затем для их сочетаний (см.п.2.3.2.) для первой записи дало значение $C_{\max}=35$, для второй - $C_{\max}=32$, для третьей - $C_{\max}=28$, что показало неэффективность модели нарушителя даже для одного идентификатора, не говоря уже о полной строке идентификаторов.

Поскольку количество совпадений $S_{\max}$ хотя и медленно, но снижается, то возможно, что при наличии у нарушителя знакомых записей в количестве гораздо большем, чем 5 (параметр данной модели), задача построения таблицы смещений будет решена полностью.

Для проверки этой гипотезы авторы сняли ограничение количества известных записей и частично сняли ограничение на разработку программного обеспечения. В остальной концепция модели была сохранена и использована для обучения нейронной сети скомпанованной на базе стандартной библиотеки для распознавания бит в строке. Параметрами нейронной сети были размер и количество полносвязных слоев. Роль обучающих записей выполняли записи, известные нарушителю. Количество нейронов выходного слоя - 64 (по количеству бит в идентификаторе). количество промежуточных слоев варьировалось от 2 до 4. Результаты эксперимента показаны на рис. 4.

Рис. 4. Количество шагов (записей)

Заданная точность распознавания 0,95 была быстрее всего достигнута в двухслой-

ной архитектуре сети при использовании около 500 обучающих записей.

Данные результаты можно рассматривать в качестве дополнительного подтверждения неэффективности применения модели нарушителя при перестановке бит в строке.

3.3. Перемешивание полей внутри группы записей.

Эксперимент, проведенный для группы из 511 записей и 6 идентификаторов, показал, что наиболее вероятно повторение значения только одного из идентификаторов, например, для отчества $M_{\text{повт}} = 0,3$. А вероятность повторения отчества при использовании второй записи уже будет равна 0,005. Т.е. ее использование приведет к решению задачи однозначно.

3.4. Перемешивание символов внутри группы записей.

Для использования подхода, описанного в п.2.3.1., необходимо соблюдение условия $m^2 < N$, где $N=511$, а значение m зависит от структуры таблицы и определяется форматом полей идентификаторов в базе, а именно: поле фамилии имеет длину 15 символов, которые в повернутой таблице займут 15 первых строк, причем в 3х первых вообще не будет пробелов (нет фамилий длиной менее 3х символов). И хотя наиболее вероятная длина фамилии - 8 символов, в 8й строке этой таблице наиболее вероятно будет всего 231 символ и 280 пробелов. Т.е. в каждой отдельно взятой строке вероятное количество пробелов, а также совпадений символов m будет различным. Например, в первых 3х строках символ «а» повторится 62 раза, символ «б» - 7 раз, и наиболее вероятна ситуация, когда все 511 мест будут заполнены различными, но повторяющимися символами. Расчеты показывают, что условие $m^2 < N$ соблюдается только начиная с 7й строки. Для остальных идентификаторов условия аналогичные.

Таким образом, первые 6 строк фамилии должны рассчитываться по усложненному методу, как это показано в п.3.2. для перемешивания бит в строке. Поскольку длина строки символов в данном случае гораздо больше, чем использовалось в расчетах для перемешивания бит, можно сделать однозначный вывод о неэффективности модели нарушителя для данного варианта метода перемешивания.

4. Заключение

В данной работе рассматривались варианты реализации искажающих методов обе-

зличивания персональных данных с учетом имеющегося опыта их внедрения и возможности их взлома с помощью модели нарушителя. Опыт внедрения искажающих методов обезличивания явно недостаточен для серьезного анализа. Тем не менее, отсутствие методик расчета эффективности является причиной недоверия разработчиков к таким методам, а также причиной применения сложных алгоритмов, отличающихся для различных частей базы данных, что значительно снижает производительность обработки и ограничивает их применение базами малого размера.

В данной работе авторы рассматривают простые алгоритмы, имеющие достаточную производительность обработки баз очень большого объема. Именно для проверки эффективности таких алгоритмов разработана модель нарушителя, использование которой также было продемонстрировано.

Таким образом, применение модели нарушителя, будучи эффективным для одних методов обезличивания (перестановка символов в строке идентификаторов, перемешивание полей в группе записей) и неэффективным для других (перестановка бит в строке идентификаторов, перемешивание символов в группе записей), показало неэффективность либо эффективность соответствующих методов обезличивания.

Использование различных параметров при моделировании действий нарушителя (например, количество известных нарушителя записей) позволяет принять для них нормативные значения и сформировать соответствующую нормативную базу.

Статья выполнена при поддержке Правительства РФ (Постановление №211 от 16.03.2013 г.), соглашение № 02.A03.21.0011.

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. Кучин И.Ю. Обработка баз данных с персонифицированной информацией для задач обезличивания и поиска закономерностей / И.Ю. Кучин // Диссертация ктн. – Астрахань: Издательство Астраханского гос. технического ун-та, 2012. – 132 с.
4. Воронин В.В. Защита персональных данных в информационных системах методом обезличивания / В.В. Воронин, Н.Л. Нехай // Информационные технологии XXI века: сборник научных трудов. – Хабаровск: Издательство Тихоокеанского гос. ун-та, 2017. – С. 479–483.
5. Карпова И.П. О реализации метода обезличивания персональных данных / И.П. Карпова // Вестник компьютерных и информационных технологий. – Москва: Издательский дом «Спектр», 2013. – № 6, – С. 56–60.
6. Бондаренко К.О. Универсальный быстродействующий алгоритм процедур обезличивания данных / К.О. Бондаренко, В.А. Козлов // Известия ЮФУ. Технические науки. – Ростов-на-Дону: Издательство Южного фед. ун-та, 2015. – № 11 (172). – С. 130–142.
7. Виленкин Н.Я. Глава III. Комбинаторика кортежей и множеств. Размещения с повторениями/ Н.Я. Виленкин // Популярная комбинаторика. — М.: Наука, 1975. — С. 80. — 208.
8. Мищенко Е.Ю. Вероятность идентификации в базе персональных данных: Выбор идентифицирующих атрибутов / Е.Ю. Мищенко // Безопасность информационного пространства: сборник трудов XVIII Всероссийской научно-практической конференции студентов, аспирантов и молодых ученых. – Магнитогорск: Изд-во Магнитогорск. гос. техн. ун-та им. Г.И. Носова, 2019. – С. 206-209.

References

1. Federal'nyy zakon ot 27.07.2006 «O personal'nykh dannykh» № 152-FZ [Elektronnyy resurs]. – Rezhim dostupa: http://www.consultant.ru/document/cons_doc_LAW_61801/, svobodnyy (data obrashcheniya: 14.05.2018).
2. Prikaz Federal'noy sluzhby po nadzoru v sfere svyazi, informatsionnykh tekhnologiy i massovykh kommunikatsiy ot 05.09.2013 № 996 «Ob utverzhdenii trebovaniy i metodov po obezlichivaniyu personal'nykh dannykh» [Elektronnyy recurs]. – Rezhim dostupa: <http://54.rkn.gov.ru/protection/acts/p13580/>, svobodnyy (data obrashcheniya: 14.05.2018).

3. Kuchin I.YU. Obrabotka baz dannykh s personifitsirovannoy informatsiyey dlya zadach obezlichivaniya i poiska zakonnomernostey / I.YU. Kuchin // Dissertatsiya ktn. – Astrakhan': Izdatel'stvo Astrakhanskogo gos. tekhnicheskogo un-ta, 2012. – 132 s.

4. Voronin V.V. Zashchita personal'nykh dannykh v informatsionnykh sistemakh metodom obezlichivaniya / V.V. Voronin, N.L. Nekhay // Informatsionnyye tekhnologii XXI veka: sbornik nauchnykh trudov. – Khabarovsk: Izdatel'stvo Tikhookeanskogo gos. un-ta, 2017. – S. 479–483.

5. Karpova I.P. O realizatsii metoda obezlichivaniya personal'nykh dannykh / I.P. Karpova // Vestnik komp'yuternykh i informatsionnykh tekhnologiy. – Moskva: Izdatel'skiy dom «Spektr», 2013. – № 6, – S. 56–60.

6. Bondarenko K.O. Universal'nyy bystrodeystviyushchiy algoritm protsedur obezlichivaniya dannykh / K.O. Bondarenko, V.A. Kozlov // Izvestiya YUFU. Tekhnicheskoye nauki. – Rostov-na-Donu: Izdatel'stvo Yuzhnogo fed. un-ta, 2015. – № 11 (172). – S. 130–142.

7. Vilenkin N.YA. Glava III. Kombinatorika kortezhey i mnozhestv. Razmeshcheniya s povtorenyami / N.YA. Vilenkin // Populyarnaya kombinatorika. — M.: Nauka, 1975. — S. 80. — 208.

8. Mishchenko Ye.YU. Veroyatnost' identifikatsii v baze personal'nykh dannykh: Vybory identifikatsionnykh atributov / Ye.YU. Mishchenko // Bezopasnost' informatsionnogo prostranstva: sbornik trudov XVIII Vserossiyskoy nauchno-prakticheskoy konferentsii studentov, aspirantov i molodykh uchenykh. – Magnitogorsk: Izd-vo Magnitogorsk. gos. tekhn. un-ta im. G.I. Nosova, 2019. – S. 206–209.

МИЩЕНКО Евгений Юрьевич, старший преподаватель кафедры защиты информации, Южно-Уральский государственный университет (национальный исследовательский университет). 454080, Россия, г. Челябинск, пр. Ленина, 76. E-mail: mishchenkoei@susu.ru

СОКОЛОВ Александр Николаевич, кандидат технических наук, доцент, заведующий кафедрой защиты информации ФГАОУ ВО "Южно-Уральский государственный университет (национальный исследовательский университет)". Россия, 454080 Челябинск, проспект Ленина, 76. E-mail: sokolovan@susu.ru

MISHCHENKO Evgeniy, Senior Lecturer of Department of Information Security, South Ural State University (national research university). 76, Lenina prosp., Chelyabinsk, Russia, 454080. Email: mishchenkoei@susu.ru

SOKOLOV Alexander, Ph.D., Associate Professor, Head of Department of Information Security, Federal State Autonomous Educational Institution of Higher Education "South Ural State University (national research university)". 76, Lenin prospekt, Chelyabinsk, Russia, 454080. E-mail: sokolovan@susu.ru