ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

кандидата технических наук, доцента Спевакова Александра Геннадьевича на диссертационную работу Мищенко Евгения Юрьевича на тему «Моделирование процессов обезличивания персональных данных и оценка эффективности используемых методов на основе модели нарушителя», представленной на соискание ученой степени кандидата технических наук по специальности 2.3.6. Методы и системы защиты информации, информационная безопасность

Актуальность темы исследования

Согласно требованиям нормативных актов, операторы информационных систем персональных данных (ИСПДн) обязаны применять меры по защите персональных данных (ПД), обрабатываемых и хранимых в базах данных (БД), входящих в состав ИСПДн, в том числе в научных, статистических и прочих целях, предполагающих предварительное обезличивание ПД.

Под обезличиванием персональных данных понимаются действия, в результате которых становится невозможным без использования дополнительной информации определить принадлежность ПД конкретному субъекту персональных данных.

В РФ процесс обезличивания персональных данных регламентирован Приказом Роскомнадзора, в соответствии с которым, для практического применения, рекомендуется использовать следующие методы обезличивания:

- введение идентификаторов – замена части сведений идентификаторами с созданием таблицы соответствия идентификаторов исходным данным;
- изменение состава или семантики – изменение состава или семантики персональных данных путем замены результатами статистической обработки, обобщения или удаления части сведений;
- декомпозиция – разбиение множества персональных данных на несколько частей с последующим раздельным хранением подмножеств;
- перемешивание – перестановка отдельных записей, а также групп записей в массиве персональных данных.

Следует отметить, что решения задачи обезличивания ПД вышеуказанными методами, имеют существенные недостатки. Так, например, отсутствовали обоснование выбора атрибутов и оценка эффективности алгоритмов обезличивания, после обработки оставалась возможность установить принадлежность субъекта ПД к конкретной ИСПДн.

Актуальность темы диссертационного исследования обусловлена необходимостью решения проблем, возникающих при реализации методов обезличивания в ИСПДн, к которым относятся:

– отсутствие методики обоснования выбора методов обезличивания и настройки их параметров в зависимости от свойств базы ПД;

– отсутствие методики количественной оценки эффективности методов обезличивания ПД.

Научная новизна полученных результатов

В диссертационной работе получены следующие новые научно обоснованные результаты:

1. Разработана математическая модель идентификации ФЛ, отличающаяся применением количественных оценок вероятности идентификации по атрибуту в целом, и сформулирован критерий необходимости обезличивания ПД по любым идентификаторам или их совокупности при любом объеме БД.

2. Разработана функциональная модель нарушителя, реализующая итерационный алгоритм деобезличивания ПД для методов обезличивания, основанных на искажении ПД, и отличающаяся применением количественных оценок эффективности методов обезличивания.

3. Разработана функциональная схема передачи информации между базами ПД, разделенными методом введения идентификаторов, отличающаяся применением внешнего носителя идентификационной информации и реализующая безопасную передачу информации между таблицей идентификаторов и обезличенными ПД.

Обоснованность и достоверность научных положений, сформулированных в диссертации

Обоснованность и достоверность научных результатов исследования достигнута благодаря корректной постановке цели и задач исследования, адекватному задачам исследования выбору математического аппарата и результатами экспериментальной апробации предложенных моделей (применимостью разработанных критериев для баз ПД различного объема).

Результаты исследования опубликованы в 13 научных работах, 6 из которых в ведущих рецензируемых журналах, рекомендованных ВАК РФ и Аттестационным советом УрФУ. Имеется 1 патент на полезную модель.

Практическая значимость результатов диссертации

Практическая значимость заключается в:

1. возможности применения критерия необходимости обезличивания ПД для обоснования выбора обезличиваемых атрибутов ФЛ;

2. возможности использования регулирующими органами показателя вероятности идентификации по атрибуту ФЛ для создания нормативной базы параметров обезличивания

ПД;

3. возможности применения функциональной схемы передачи информации между частями базы ПД, разделенными методом введения идентификаторов, для построения системы защиты ИСПДн.

Оценка содержания диссертации и её оформления

Диссертационная работа содержит 141 страницу основного текста (всего 165 с.), 58 рисунков (без приложений) и 22 таблицы (без приложений). Состоит из введения, четырех глав, заключения, списка сокращений и условных обозначений, списка литературы из 93 наименований, 4 приложений.

В первой главе подробно рассмотрено состояние предметной области, сделан акцент на результатах, полученных как в работах отечественных, так и зарубежных учёных, описывающих существующие реализации методов обезличивания, их технологические особенности и возможности оценки эффективности. Автор выделяет две группы методов обезличивания ПД – основанные на разделении БД и основанные на внесении обратимых искажений, систематизировано рассматривает сходство и различие проблем реализации различных методов обезличивания. На основе анализа состояния предметной области автор обосновывает необходимость разработки моделей процесса обезличивания ПД ФЛ, методики оценки эффективности реализации методов обезличивания и схемы защищенной передачи информации между разделенными частями БД, формулирует цель и задачи исследования.

Вторая глава посвящена разработке модели идентификации ФЛ и модели нарушителя безопасности ПД, на основе взаимодействия которых сформулирован критерий необходимости обезличивания ПД ФЛ по атрибутам. Модель идентификации ФЛ основана на использовании понятия вероятности идентификации по атрибуту в целом, значение которого определяется путем применения методов частотного анализа и параметрической идентификации к эмпирическим последовательностям характеристик атрибутов ФЛ и их сочетаний, с учетом их семантики и объемов БД. Для учета возможностей нарушителя предложена методика построения общей модели нарушителя безопасности ПД, не зависящая от метода обезличивания, определены ее параметры, влияющие на критерий необходимости обезличивания. Согласно этому критерию, в работе рассмотрена необходимость обезличивания для восьми различных атрибутов ФЛ и некоторых их сочетаний, а также для десяти случайных выборок БД различного объема. Достоверность результатов применения данного критерия подтверждена для двух БД большого объема, имеющих различную структуру.

Третья глава посвящена разработке модели нарушителя, реализующей алгоритм

деобезличивания, и методики оценки эффективности обезличивания для методов обезличивания, основанных на искажении ПД. Одним из параметров модели нарушителя является количество записей ФЛ, ПД которых достоверно известны нарушителю. В качестве алгоритмов искажения рассмотрены четыре алгоритма, основанные на перемещении элементов (символов и битов): два для метода изменения состава или семантики и два для метода перемешивания, для каждого из которых автором предложен алгоритм деобезличивания. Применив в алгоритмах деобезличивания методы комбинаторного анализа, автор произвел количественную оценку возможности достижения нарушителем результата, заданного параметром эффективности, что позволило обосновать эффективность либо неэффективность рассмотренных алгоритмов обезличивания.

В четвертой главе предложена функциональная схема передачи информации между частями базы ПД, разделенными методом введения идентификаторов. В представленной схеме для связи таблицы, содержащей чувствительные атрибуты, с обезличенной частью БД используется внешний идентификатор. В основу схемы положена полезная модель, на которую автором получен патент. Эффективность результатов внедрения схемы оценена в рамках рассмотренной в работе модели нарушителя.

Замечания и вопросы по работе

Тем не менее, следует отметить некоторые замечания и вопросы по содержанию автореферата и диссертации:

1. При формулировании критерия необходимости обезличивания (раздел 2.2 диссертации) важная роль отводится параметру модели нарушителя U, связанному с компетенцией нарушителя. Предлагается значение параметра $U = 20$ и предполагается, что оно будет нормироваться регулятором. Какое нормированное значение имеется ввиду – максимальное, среднее или иное?

2. При построении диаграмм частот для некоторых атрибутов (например, в разделе 2.3.3 диссертации) наблюдается отклонение от монотонности в области максимальных значений q (количества записей, имеющих одинаковое значение атрибута) – так называемый эффект «размытия хвоста». Почему для устранения последствий этого эффекта автор использует метод изменения размера интервалов, а не метод их отсечения?

3. При определении критерия обезличивания графики распределения вероятности для различных атрибутов, как для подлежащих обезличиванию (рис. 4 диссертации), так и для не подлежащих обезличиванию (рис. 9 диссертации), показывают избыточность вероятности идентификации $W_{норм}$ относительно возможностей нарушителя U. В связи с этим возникают следующие вопросы:

1) Является это отношение обязательным, характерным или случайным?

2) Какое практическое значение имеет нарушение этого отношения для атрибута «дата рождения (рис. 30 диссертации)?

4. При рассмотрении модели нарушителя для методов обезличивания, основанных на искажении идентификаторов (стр. 102 диссертации), для алгоритма деобезличивания задается значение параметра $G = 5$ (нарушителю известны персональные данные пяти физических лиц в обезличенной базе). Какова будет итоговая оценка эффективности метода обезличивания одной и той же базы при различных значениях G?

5. На рис. 43 в тексте диссертации приведен алгоритм поиска смещений символов в обезличенной строке идентификаторов для количества известных нарушителю записей в базе G = 3. Каким образом на последнем шаге (элемент сравнения «Все смещения определены?») учитывается параметр компетенции нарушителя U?

Сделанные замечания имеют дискуссионный характер и не снижают научной ценности работы.

Заключение по работе

Исследовательская работа изложена грамотным научно-техническим языком, в полной мере отвечает требованиям по актуальности, научной новизне, практической значимости, личному вкладу автора, отражению результатов в публикациях, а также полностью соответствует п. 9 Положения о присуждении ученых степеней в УрФУ и специальности 2.3.6. Методы и системы защиты информации, информационная безопасность. Автор диссертации Мищенко Евгений Юрьевич заслуживает присуждения ему учёной степени кандидата технических наук по специальности 2.3.6. Методы и системы защиты информации, информационная безопасность.

Официальный оппонент:

Кандидат технических наук, доцент
Доцент кафедры информационной безопасности
ФГАОУ ВО «Московский политехнический университет»
e-mail: aspev@yandex.ru
Адрес: ул. Б. Семёновская, 38, Москва, 107023

18.01.2023 Спеваков Александр Геннадьевич
(подпись) (дата)

Подпись к.т.н., доцента
Спевакова А.Г. заверяю СПЕЦИАЛИСТ ПО КАДРОВОМУ ДЕЛОПРОИЗВОДСТВУ БИРЮКОВА И.
(подпись) (дата)