УДК 004.891.3
DOI: 10.25559/SITITO.18.202203.666-679

Оригинальная статья

Методика оценки риска от разглашения конфиденциальной информации в источниках данных с использованием интеллектуального анализа данных

А. И. Шаброва, А. А. Теренин*, Н. Г. Бабак
ПАО «Сбербанк России», г. Москва, Российская Федерация
Адрес: 117312, Российская Федерация, г. Москва, ул. Вавилова, д. 19
* Terenin.A.Alek@sberbank.ru

Аннотация

На сегодняшний день, степень развития методов и средств, позволяющих оценить уровень риска от распространения конфиденциальной информации в источниках, не предназначенных для этого, достаточно низкая. В современном мире многие коммерческие организации собирают информацию о своих клиентах, а также хранят и обрабатывают информацию о собственной деятельности и средствах достижения финансовых результатов. Проблема заключается в том, что отсутствует единая методика оценки остаточного риска при публикации чувствительной информации в источниках, не предназначенных для этого. Также отсутствует система для регулярной оценки данного типа риска. Целью исследования является проверка гипотезы о возможности и необходимости регулярного мониторинга источников данных с целью выявления наличия чувствительной информации и её защиты путем создания методики оценки рисков от разглашения конфиденциальной информации. Новизна работы заключена в разработке авторского алгоритма оценки рисков от распространения конфиденциальной информации и построении математической модели, позволяющей произвести количественную оценку рисков, предложены варианты определения вероятностей наступления событий и способ задания и использования шкалы, основанной на экспертных оценках. Для достижения поставленной в исследовании цели используются общенаучные методы в рамках сравнительного и статистического анализа, а также экспертные оценки и графическая интерпретация полученных в ходе исследования результатов. В качестве предлагаемого решения проблемы представлена авторская модификация трехфакторной модели оценки рисков и адаптированный подход к достижению допустимого уровня риска от разглашения конфиденциальной информации. В результате проведенного анализа оценен риск от разглашения чувствительной информации, на примере открытых источников информации, при помощи предложенного алгоритма оценки и разработанный математической модели, а также выявлены проблемные места, определена шкала риска для рассматриваемых в исследовании источников информации. Еще раз подтверждена необходимость в развитии систем, позволяющих оценивать уровни риска от разглашения чувствительной информации, развитии методов и подходов к алгоритмам выявления и предотвращения таких разглашений.

Ключевые слова: оценка рисков информационной безопасности, конфиденциальная информация, интеллектуальный анализ данных, защита данных, искусственный интеллект, машинное обучение, кибербезопасность, персональные данные, банковская тайна, коммерческая тайна

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Для цитирования: Шаброва А. И., Теренин А. А., Бабак Н. Г. Методика оценки риска от разглашения конфиденциальной информации в источниках данных с использованием интеллектуального анализа данных // Современные информационные технологии и ИТ-образование. 2022. Т. 18, № 3. С. 666-679. doi: https://doi.org/10.25559/SITITO.18.202203.666-679

Original article

Methodology for Risk Assessment from Confidential Information Disclosure in Data Sources Using Data Mining

A. I. Shabrova, A. A. Terenin*, N. G. Babak
PJSC "Sberbank of Russia", Moscow, Russian Federation
Address: 19 Vavilov St., Moscow 117312, Russian Federation
* Terenin.A.Alek@sberbank.ru

Abstract

At the moment, the low level of development of methods and tools for assessing the level of risk from the dissemination of confidential information in sources in which such data should not be. In the modern world, many commercial organizations collect information about their customers, store and process information about their own activities and means of achieving financial results. The problem is that there is no single methodology for assessing the risk associated with storing confidential information in sources that should not contain such data. There is also no system for regular assessment of this type of risk. The purpose of the study is to test the hypotheses about the possibility and necessity of regular monitoring of data sources in order to identify confidential information and protect it using the developed methodology for assessing the risks of disclosing confidential information. The novelty of the study lies in the development of the author's algorithm for assessing risks from the dissemination of confidential information and the construction of a mathematical model that allows for a quantitative assessment of risks, options for determining the probabilities of occurrence of events and a methodology for establishing and using a scale based on expert assessments. To achieve the goal set in the study, general scientific methods are used in the framework of comparative and statistical analysis, as well as expert assessments and graphical interpretation of the results obtained during the study. The author's modification of the three-factor risk assessment model and an adapted approach to achieving an acceptable level of risk from the disclosure of confidential information are presented as a solution to the problem. As a result of the analysis, the risk of disclosing confidential information was assessed, problem areas were identified using the example of open sources of information, and a scale of riskiness of sources was determined. Once again, the need to develop systems that allow assessing the levels of risk from the disclosure of confidential information, the development of methods and approaches to algorithms for detecting and preventing such disclosures has been confirmed.

Keywords: information security risk assessment, confidential information, data mining, data protection, artificial intelligence, machine learning, cybersecurity, personal data, bank secrecy, trade secret

Conflict of interests: The authors declare no conflict of interest.

For citation: Shabrova A.I., Terenin A.A., Babak N.G. Methodology for Risk Assessment from Confidential Information Disclosure in Data Sources Using Data Mining. *Modern Information Technologies and IT-Education.* 2022;18(3):666-679. doi: https://doi.org/10.25559/SITITO.18.202203.666-679

Введение

В последнее время организации все больше ориентированы на сбор большого количества данных о своих клиентах, сотрудниках, подрядчиках и других агентах того рынка, на котором занимаются своей производственной деятельностью. В связи с этим, актуальным становится вопрос защиты этих данных и сохранения их недоступности сторонним агентам.

Для того чтобы защищать информацию, необходимо четко определять её категорию и уровень доступности. В работе исследуются следующие типы конфиденциальной (чувствительной) информации: персональные данные, банковская тайна и коммерческая тайна.

Также, необходимо определить перечень исследуемых источников информации, убедиться в их разнородности, для более широкого охвата источников, в зависимости от целей, которые преследуют её обладатели и пользователи.

Актуальность исследуемой темы подтверждается следующим:

1. Отсутствие методологии оценки данного типа риска.
2. Отсутствие систем для регулярной оценки данного типа риска.
3. Недостаточный уровень осведомленности сотрудников относительно вопросов защиты информации внутри организаций.

Цель исследования

Проверка гипотезы о необходимости регулярного мониторинга источников информации с целью обнаружения и предотвращения разглашения конфиденциальной информации, путем создания алгоритма оценки рисков от её распространения.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Выполнить аналитический обзор существующих областей защиты данных и методов оценки рисков.
2. Сформировать требования к системе оценки рисков инцидентов разглашения конфиденциальной информации в открытых источниках.
3. Разработать модель интеллектуальной обработки информации из открытых источников.
4. Выдвинуть и проверить гипотезы о необходимости мониторинга источников информации.
5. Разработать методику оценки рисков распространения конфиденциальной информации в источниках информации, не предназначенных для этого.
6. Получить результаты применения разработанного алгоритма оценки рисков для банковской сферы.

В рамках данного исследования разработаны:

1. Алгоритм оценки рисков распространения конфиденциальной информации в открытых источниках.
2. Подход к интеллектуальной обработки информации, в рамках решаемой задачи.
3. Методика оценки рисков распространения конфиденциальной информации в источниках информации.

Практическая значимость результатов исследования заключается в возможности использования разработанного процесса (от сбора и предобработки информации до получения количественной оценки рисков) для идентификации инцидентов и оценки рисков от разглашения конфиденциальной информации в открытых источниках в организациях, которые владеют информацией о своих клиентах, других компаниях и внутренними разработками, позволяющими иметь конкурентное преимущество на рынке. Результаты данного анализа служат аргументами в пользу развития и модернизации мер по защите конфиденциальной информации.

Основная часть

Основную часть исследования необходимо начать с главных определений рассматриваемой предметной области.

Согласно федеральному закону от 27 июля 2006 г. № 152-ФЗ, персональные данные – любая информация, относящаяся к прямо или косвенно определенному, или определяемому физическому лицу (субъекту персональных данных). Согласно этому документу, такой тип данных может обрабатываться только для личных и семейных нужд, при этом не нарушая права субъекта этих данных, а также для нужд архивного фонда Российской Федерации и в других случаях, предусмотренных законом[1]. Что подтверждает, что несогласованная обработка персональных данных является инцидентом кибербезопасности. Любая обработка данных должна осуществляться на законной и справедливой основе, при этом обработка данных возможна только для достижения определенных целей.

Согласно федеральному закону от 29 июня 2012 г. № 97-ФЗ, банковская тайна – информация об операциях, о счетах и вкладах своих клиентов и корреспондентов. За разглашение банковской тайны организации и ее руководители несут ответственность, включая возмещение нанесенного ущерба, в порядке, установленном федеральным законом[2].

Согласно федеральному закону от 29 июля 2004 г. № 98-ФЗ, коммерческая тайна – режим конфиденциальности информации, позволяющий ее обладателю при существующих или возможных обстоятельствах увеличить доходы, избежать неоправданных расходов, сохранить положение на рынке товаров, работ, услуг или получить иную коммерческую выгоду[3].

Актуальность исследуемого вопроса подтверждается интересом к защите данных и другим вопросам кибербезопасности. Существуют исследования, доказывающие необходимость

[1] О персональных данных: федер. закон от 27 июля 2006 г. № 152-ФЗ (с изменениями на 14 июля 2022 года) [Электронный ресурс]. URL: http://www.consultant.ru/document/cons_doc_LAW_61801 (дата обращения: 07.09.2022).

[2] О внесении изменений в часть первую и часть вторую Налогового кодекса Российской Федерации и статью 26 Федерального закона «О банках и банковской деятельности»: федер. закон от 29 июня 2012 г. № 97-ФЗ (ред. от 02.06.2016) [Электронный ресурс]. URL: http://www.consultant.ru/document/cons_doc_LAW_131891/30b3f8c55f65557c253227a65b908cc075ce114a/#dst100175 (дата обращения: 07.09.2022); О банках и банковской деятельности: федер. закон от 02 декабря 1990 г. № 395-1 (последняя редакция) [Электронный ресурс]. URL: http://www.consultant.ru/document/cons_doc_LAW_5842 (дата обращения: 07.09.2022).

[3] О коммерческой тайне: федер. закон от 29 июля 2004 г. № 98-ФЗ (последняя редакция) [Электронный ресурс]. URL: http://www.consultant.ru/document/cons_doc_LAW_48699 (дата обращения: 07.09.2022).

повышать осведомленность сотрудников в области кибербезопасности [1]. Другие исследования направлены на изучение вопроса сохранности конфиденциальных данных, в частности персональных данных, организациями. Поднимается проблема использования мобильных устройств в работе сотрудников и несения ответственности за правильную эксплуатацию всех программных средств [2]. Также, в некоторых работах подчеркнута привлекательность общего пользования данными организации с зависимыми компаниями (дочерними, зависимыми и партнёрскими), при этом отмечая необходимость защиты этих данных с точки зрения доступности, посредством защищенных механизмов и систем [3, 4].

Помимо этого, в ряде исследований поднимается вопрос инвестирования в системы по обнаружению инцидентов кибербезопасности, а не предотвращению. Это интуитивно понятно, поскольку ресурсы, вложенные в обнаружение, социально более полезны, чем те, которые направлены на сокращения потерь после того, как атаки произошли или остались незамеченными [5, 6]. По результатам анализа авторы пришли к выводу, что оптимальное решение относительно инвестирования в кибербезопасность – это сочетание портфеля мер противодействия (обнаружение, предотвращение или сдерживание). Многие источники подчеркивают, что развитие новых технологий и сервисов приносят пользу как предприятиям, так и потребителям [7], они также создают серьезные риски конфиденциальности. Что влияет на отношение людей к коммерческим организациям, которые собирают данные для своих производственных целей. В таком случае, если не применять методы защиты данных, отсутствие доверия может замедлить развитие инноваций, активное использование и внедрение новых технологий, что в свою очередь приведет к упущению возможностей для бизнеса [8, 9].

В целом, повышение осведомленности сотрудников и их обучение новым и планируемым изменениям в области защиты конфиденциальной информации – это ключ к минимизации нарушений, связанных с разглашением или неправомерным использованием [10].

Необходимо повышать уровень защищенности этих данных для сторонних агентов, четко понимая последствия, которые могут возникнуть при нарушении конфиденциальности данных клиентов. Например, важным аспектом в использовании персональных данных клиентов является согласие этого клиента на хранение и обработку его персональных данных, с четким указанием целей анализа его данных. В связи с этим важно четко определить направления аналитики и защищенность от утечек или перехвата сторонними организациями. Также, некоторые исследователи в своих работах подчеркивают недостаточность технических средств для идентификации нарушений информационной безопасности и запуска мер по их предотвращению. Основная рекомендация авторов – создание единой системы с постоянным мониторингом сохранения конфиденциальности и единого подхода, хотя бы внутри одной организации и взаимосвязанными с ней компаниями (если такая структура имеет место быть).

Существуют работы, направленные на расширение точек зрения в исследовании вопроса защиты конфиденциальных данных. Например, поднимается вопрос о групповой конфиденциальности данных [11], в том смысле, что при аналитике многие организации посредством алгоритмов анализа больших данных группируют клиентов по определенным признакам. В связи с этим важно защищать не только персональные данные одного клиента, но и всей группы. Поднимается вопрос недостаточной защищенности групповой информации. Основной аргумент против использования коллективных данных – это борьба с ценовой дискриминацией. Данный взгляд подтверждает, что в вопросе защиты персональных данных необходимо выходить за пределы устоявшихся правил и смотреть на вопрос более глобально и изучать влияние нарушений взаимосвязанных с персональными и другими видами данных.

Также, например, при анализе открытых источников, существует актуальная и сложная задача реализации разделения тех пользователей, которые сами ведут себя неосторожно, раскрывая свои данные, от тех, чья информация раскрывается без их ведома. В рамках исследования выявлено, что пользователи положительно реагируют на стимулы социальных сетей к защите данных. Сейчас, некоторые социальные сети позволяют пользователям отключить возможность сбора пользовательских данных сторонними веб-сайтами или приложениями. Результатом исследования стал вывод, что пользователи склонны к тому, чтобы самостоятельно управлять своими данными, сохраняя возможность их конфиденциальности. Также важным результатом данной работы является перечень поведенческих мер, которые пользователь принимает при идентификации инцидента о разглашении информации о нем: блокировка, удаление, игнорирование.

Некоторые исследователи показывают, что нарушение политик конфиденциальности персональных данных ведет к экономическим потерям для потребителей (например, на ценовой дискриминации), не говоря уже о более серьезных последствиях, когда происходит «кража личности». Аргументация в пользу систем, позволяющих предотвращать утечки информации, заключается в том, что потеря или кража конфиденциальных данных ведет не только к прямым финансовым убыткам, но и к снижению доверия со стороны клиентов, партнеров и инвесторов[4]. Любая утечка данных приводит к повышенному интересу со стороны регулирующих органов и СМИ. Это увеличивает риски репутационной и финансовой ответственности за нарушение отраслевых стандартов и законодательных актов, регулирующих защиту персональных данных и другой конфиденциальной информации [12-15].

Также, во многих недавних отчетах «Гартнер»[5], рекомендуется планировать стратегические направления в области управления данными, подчеркивается важность выявления конфиденциальных данных во всех типах информационных активов,

[4] Bridal O. Named-entity recognition with BERT for anonymization of medical records : Bachelor's thesis. Linköping: Linköping University, 2021. [Электронный ресурс]. URL: https://www.diva-portal.org/smash/get/diva2:1566701/FULLTEXT01.pdf (дата обращения: 07.09.2022).

[5] Cearley D. W., Burke B., Searle S., Walke M. J. Top 10 Strategic Technology Trends for 2018 [Электронный ресурс] // Gartner. 03.10.2017. URL: https://www.gartner.com/ngw/globalassets/en/information-technology/documents/top-10-strategic-technology-trends-for-2018.pdf. (дата обращения: 07.09.2022).

включая социальные сети, новостные отчеты, внутренние системы и др[6].

Далее представлен анализ существующих методологий, с помощью которых можно оценить риски.

На сегодняшний момент не удалось установить единого подхода к количественным и качественным оценкам рисков информационной безопасности, связанных с разглашением конфиденциальной информации. Существуют работы, затрагивающие необходимость в подобной оценке, а также исследования, рассматривающие подходы к оценке рисков кибербезопасности, но в отношении атак, продажи больших объемов инсайдерской и клиентской информации организаций [16].

Также, как и в анализе предметной области, следует начать с основных определений[7].

Риск – сочетание вероятности события и его последствий (результатов событий, которые могут быть выражены качественно или количественно). Риск информационной безопасности – потенциальная возможность использования уязвимостей актива или группы активов конкретной угрозой информационной безопасности для причинения ущерба организации[8].

Актив – все, что имеет ценность для организации и находится в ее распоряжении или то, что обладает ценностью или полезностью для организации, ее бизнес-операций и их непрерывности, и поэтому нуждается в защите, которая позволит обеспечить корректное выполнение бизнес-операций и непрерывность бизнеса. Ресурс – актив организации, который используется или потребляется в процессе выполнения некоторой деятельности. Информация в современном мире – это ресурс.

Угроза информационной безопасности – совокупность условий и факторов, создающих потенциальную или реально существующую опасность нарушения свойств информационной безопасности – конфиденциальности, доступности и/или целостности информации/информационных активов организации[9].

Оценка риска – основанная на результатах анализа риска процедура проверки, устанавливающая, не превышен ли допустимый риск. Допустимый риск – который считается в данной ситуации приемлемым при существующих общественных ценностях. Анализ риска – систематическое использование информации для выявления опасности и количественной оценки риска[10].

В случае, если при оценке рисков, составляющие элементы являются категориальными значениями, то необходимо создать таблицу с оценкой и качественным описанием уязвимостей.

Для оценки риска разглашения конфиденциальной информации предлагается методология, основанная на ГОСТ Р 51898-2002 «Аспекты безопасности. Правила включения в стандарты»[11]:

1. Идентификация рисков информационной безопасности (определение слабых мест, которые могут быть использованы источником угрозы для причинения вреда активам), которая должна включать все риски, даже если они не попадают под контроль организации (как в данном исследовании);
   - идентификация активов (определение исследуемых активов, в данной работе выбраны: персональные данные, банковская тайна и коммерческая тайна);
   - выявление возможных угроз информационной безопасности (определено, что под угрозами понимаются события, которые могут произойти в случае возникновения разглашения конфиденциальной информации: репутационные потери, финансовые потери, выраженные в штрафах от регулирующих органов, финансовые потери от снижения конкурентоспособности на рынке и утечке клиентов, нарушение законодательства, ведущее к судебному процессу и затратам на него);
   - определение существующих в организации средств управления рисками информационной безопасности, с целью идентификации слабых мест в управлении данными по итогам исследования;
   - идентификация уязвимостей (уязвимость в данном случае определена как возможность несанкционированно реализовать инцидент, с точки зрения доступности информации и выражается в информированности людей и доступности технических средств для этой реализации);
   - определении возможных последствий инцидентов разглашения информации (в данном исследовании принято, что рассматриваются последствия от угроз, указанных во втором пункте данного списка);
2. Количественная оценка рисков информационной безопасности;
   - оценка последствий от инцидентов разглашения рассматриваемого типа информации (количественная величина потерь, выраженная в денежных суммах или количестве клиентов);

[6] Варфоломеев А. А. Управление информационными рисками: учеб. пособие. М. : РУДН, 2008. 158 с.

[7] Стандарт Банка России СТО БР ИББС-1.0-2014 «Обеспечение информационной безопасности организаций банковской системы Российской Федерации. Общие положения»: принят и введен в действие распоряжением Банка России от 17 мая 2014 г. № Р-399. М. : 2014. URL: https://www.cbr.ru/Crosscut/LawActs/File/446 (дата обращения: 07.09.2022).

[8] ГОСТ Р ИСО/МЭК 27005-2010 Информационная технология. Методы и средства обеспечения безопасности. Менеджмент риска информационной безопасности: национальный стандарт РФ: издание официальное: утвержден и введен в действие Приказом Федерального агентства по техническому регулированию и метрологии от 30 ноября 2010 г. № 632-ст: введен впервые: дата введения 2011-12-01 / подготовлен ООО НПФ «Кристалл», ФГУ «ГНИИИ ПТЗИ ФСТЭК России». М. : Стандартинформ, 2019.

[9] ГОСТ Р 50922-2006 Защита информации. Основные термины и определения: национальный стандарт РФ: издание официальное: утвержден и введен в действие Приказом Федерального агентства по техническому регулированию и метрологии от 27 декабря 2006 г. № 373-ст: введен впервые: дата введения 2008-02-01 / подготовлен ФГУ «ГНИИИ ПТЗИ ФСТЭК России». М. : Стандартинформ, 2008.

[10] ГОСТ Р 51897-2021 Менеджмент риска. Термины и определения: национальный стандарт РФ: издание официальное: утвержден и введен в действие Приказом Федерального агентства по техническому регулированию и метрологии от 11 ноября 2021 г. № 1489-ст: введен впервые: дата введения 2022-03-01 / подготовлен АРМ «РусРиск». М. : Стандартинформ, 2021.

[11] ГОСТ Р 51898-2002 Аспекты безопасности. Правила включения в стандарты: национальный стандарт РФ: издание официальное: утвержден и введен в действие Постановлением Госстандарта России от 5 июня 2002 г. № 228-ст: введен впервые: дата введения 2003-01-01 / подготовлен ТК 10. М. : Стандартинформ, 2018.

- оценка вероятностей угроз (важно учитывать частоту угрозы, мотивацию к нарушению, доступность информации), также, на данном этапе определен подход к вычислению вероятностей;
- определение уровня (величины) рисков информационной безопасности (данный этап содержит разработку и утверждение модели для оценки рисков, определение факторов и их количественных выражений);

3. Оценивание рисков информационной безопасности (непосредственно, получение количественных величин уровня риска для рассматриваемого типа информации и источника, в котором реализован инцидент разглашения).

Данный подход к оценке рисков информационной безопасности, связанных с разглашением конфиденциальной информации в открытых источниках, позволяет избежать некоторых недостатков большинства современных подходов к оценке рисков информационной безопасности, а именно:

1. Избежать субъективность, которая возникает в случае определения уровня риска от реализации инцидентов, основываясь только на экспертные оценки. Разработанный подход основан на применении математического моделирования для оценки уровней риска и представляет собой основу для более эффективного управления;
2. Используются актуальные данные для получения и пояснения оценок рисков, которые регулярно обновляются и дополняются, с возможностью научно – методического прогнозирования в развитии различных сценариев угроз информационной безопасности для разных условий изменяющихся и дополняющихся факторов, влияющих на уровни риска. Помимо этого, появляется возможность отслеживать рассчитанные значения в динамике и сравнивать их с предыдущими и прогнозными величинами.

В качестве математической модели для оценки рисков информационной безопасности разработана модель, являющаяся модернизацией трехфакторной модели.

$$R = \sum_{i=1}^{4} Pr_i^{threat} \times P_i^{incidents} \quad (1),$$

$$P_i^{incidents} = Pr_i^{vul} \times P_i \quad (2),$$

$$Pr_i^{vul} \times P_i = \sum_{j=1}^{3} Pr_j^{vul} \times P_j \quad (3),$$

$$P_j = \sum_{k=1}^{5} Pr_k^{effect} \times P_k \quad (4),$$

где R – уровень риска исследуемого источника для каждого отдельного типа информации (которые заранее определены), а также для каждого заранее определенного источника для исследования, $P_i^{incidents}$ – цена потери от наступления одного из четырех событий, которые определены ниже, Pr_i^{threat} – вероятность наступления события угрозы, Pr_i^{vul} – вероятность наступления угроз с разными уровнями, P_k – потери от наступления определенных последствий при реализации угрозы, Pr_k^{effect} – вероятность наступления последствий, разных по уровню своей критичности.

Для анализа в данном исследовании выбрано несколько разнородных источников, с точки зрения аудитории авторов и типа текстовых сообщений, интересных для анализа риска разглашения конфиденциальной информации организации банковского сектора.

Примером профессионального сообщества выбран сайт Банки.ру, в котором помимо информации для потребителей банковских продуктов содержатся тематические форумы, актуальные новости о финансовых компаниях. Предполагается, что данный источник может иметь более высокий уровень риска, с точки зрения распространения коммерческой тайны и персональных данных.

В качестве другого источника выступает социальная сеть – ВКонтакте (VK). Данный ресурс позволяет пользователям создавать персональные профили и сообщества, обмениваться информацией в различных форматах и особенно популярен среди русскоязычных пользователей. Помимо этого, ресурс позволяет создавать страницы и сообщества организаций, где зарегистрированные пользователи могут проходить опросы, писать комментарии, участвовать в обсуждениях. Компании на своих страницах часто публикуют актуальные новости о себе, позволяют оставлять отзывы, задавать вопросы, писать комментарии. Помимо этого, на данном ресурсе существует множество страниц, созданных не только самими организациями, но и заинтересованными лицами, на которых часто можно встретить непредвзятое мнение экспертов и пользователей о деятельности организации, ее продуктах и мнение относительно самой организации. При этом открытость имен позволяет с большей точностью идентифицировать автора сообщения и сравнить его, например, с базой сотрудников или клиентов организации. Предполагается, что данный источник, может быть, уязвим относительно инцидентов, связанных с разглашением всех рассматриваемых типов информации: персональных данных, банковской тайны и коммерческой тайны.

В качестве источника новостей и реакции на них выбран сайт rbc.ru, данный ресурс содержит аналитические статьи о проблемах бизнеса и секторов экономики, прогнозы и анализ касательно тенденций экономики, достижения и опыт различных организаций. Часто, но не всегда, авторы пишут от своего имени, не скрывая его, что позволяет идентифицировать причастность к анализируемой или конкурирующей компании. Наиболее интересен данный источник с точки зрения разглашения коммерческой и банковской тайнах.

Количество собранных данных с источников соответствует следующим величинам: banki.ru – около 4-х тысяч записей, сайт РБК – около 2-х тысяч записей, социальная сеть vk.com – около 8 тысяч записей.

Прежде чем применять предложенную выше методику, необходимо определить допустимый уровень риска, для этого сформирована таблица 1, которая отражает зависимости от потенциальных событий, уровней уязвимости, определенных как возможность совершить разглашение конфиденциальной информации и критичности последствий для заранее определенных типов информации, также определены вероятности (шансы) наступления предложенных событий и последствий. Количественное выражение этих зависимостей представляет собой потери, определенные экспертным путем по семибалльной шкале. После определения допустимого уровня риска появляется возможность оценить эффективность контрмер.

В качестве событий – угроз для оценки ущерба выбраны:

- репутационные потери;
- финансовые потери от оплаты штрафов, выставлен-

ных государственными или надзорными органами;
- финансовые потери от утечки клиентов и разрыва партнерских отношений, что ведет к снижению конкурентоспособности;
- нарушение законодательства, которое ведет к затратам на судебные разбирательства, а также может привести к приостановке или даже прекращению основной деятельности организации.

В таблице 1 заданы возможные события – угрозы, уровни критичности.

Таблица 1. **Таблица потерь, в зависимости от типа информации, критичности и угроз**
Table 1. **Table of losses, depending on the type of information, criticality and threats**

Тип	Критичность / Событие - угроза	Уровень уязвимости	Можно пренебречь воздействием	Последствия можно легко устранить	Устранение требует умеренных затрат	Серьезные последствия	Критические последствия, останавливающие функционирование системы или организации
ПДн, БТ, КТ	Репутационные потери	Низкий	1	2	3	4	5
		Средний	2	3	4	5	6
		Высокий	3	4	5	6	7
	Финансовые потери – штрафы	Низкий	1	2	3	4	5
		Средний	2	3	4	5	6
		Высокий	3	4	5	6	7
	Финансовые потери – потеря клиентов или снижение конкурентоспособности	Низкий	1	2	3	4	5
		Средний	2	3	4	5	6
		Высокий	3	4	5	6	7
	Нарушение законодательства	Низкий	1	2	3	4	5
		Средний	2	3	4	5	6
		Высокий	3	4	5	6	7

Далее рассматриваются градации уязвимости для каждого из 4-х типов событий-угроз и каждого из 3-х видов конфиденциальной информации. Значения, по которым анализируемый текст относится к определенной степени уязвимости, заданы на основе собранных данных с использованием процентильных значений для определения пороговых значений, а также, при помощи экспертной оценки и корректировки значений, полученных на основе статистики.

Первая выделенная угроза от совершения инцидента разглашения конфиденциальной информации в открытых источниках – репутационные потери. Для данного вида угроз определено 3 градации уязвимости:

1. Низкий уровень. К данной категории относится информация, содержащая общие сведения. Для персональных данных: не более трех косвенных идентификаторов. Для банковской тайны: также, не более трех косвенных идентификаторов. Для коммерческой тайны: общие указания о внутренних разработках и перспективах развития, без указания каких-либо технических средств, сроков;
2. Средний уровень. Публикация относится к этой категории, если содержит более подробные конфиденциальные сведения. Для персональных данных: более трех косвенных идентификаторов и один прямой идентификатор субъекта информации. Для банковской тайны: более трех косвенных идентификаторов и один прямой идентификатор, а также информация о движениях средств или отношениях с другими участниками рынка. Для коммерческой тайны: указание сведений о внутренних разработках с точным указанием технологий, которые будут использованы для этих разработок, указание каких-либо фрагментов программного кода и сроков реализации;
3. Высокий уровень. Основной критерий, по которому информация относится к данной категории – широкий набор высокочувствительных данных о субъекте информации. Для персональных данных: более одного прямого идентификатора или указание номера карты. Для банковской тайны: также, более одного прямого идентификатора и информация о транзакционной активности или цели партнерских отношений. Для коммерческой тайны: подробное описание используемых средств для реализации внутренних разработок, текст внутренней нормативной и проектной документации, а также наличие полных блоков программного кода и алгоритмов, такая информация позволяет полностью реализовать работающий прототип разработанного внутри организации продукта.

Следующая выделенная угроза от совершения инцидента разглашения конфиденциальной информации в открытых источ-

никах – финансовые потери, выраженные оплатой выставленных штрафов. Для данного вида угроз также определено 3 градации уязвимости:

1. Низкий уровень. К данной категории относится информация, содержащая небольшое количество сведений о субъекте информации. Для персональных данных: не более пятидесяти инцидентов. Для банковской тайны: не более двадцати инцидентов. Для коммерческой тайны: не более десяти инцидентов;
2. Средний уровень. Источник информации относится к этой категории, если содержит большее количество конфиденциальных сведений. Для персональных данных: не более ста инцидентов. Для банковской тайны: не более пятидесяти инцидентов. Для коммерческой тайны: не более пятнадцати инцидентов;
3. Высокий уровень. Основной критерий, по которому исследуемый источник относится к данной категории – широкий набор высокочувствительных данных о субъекте информации. Для персональных данных: не более ста пятидесяти инцидентов. Для банковской тайны: не более семидесяти инцидентов. Для коммерческой тайны: не более двадцати пяти инцидентов.

Еще одна выделенная угроза от совершения инцидента разглашения конфиденциальной информации в открытых источниках – финансовые потери, выраженные утечкой клиентской базы. Для данного вида угроз аналогично предыдущим описаниям определено 3 уровня уязвимости:

1. Низкий уровень. К данной категории относится информация, содержащая небольшое количество сведений о субъекте информации, но наиболее критичной с точки зрения чувствительности информации. Для персональных данных: от одного до трех прямых идентификаторов, исключая данные карт и паспортных данных. Для банковской тайны: от одного до трех прямых идентификаторов, исключая данные расчетных счетов и движения средств по ним. Для коммерческой тайны: описание принципов работы закупленных или продаваемых программных и проектных продуктов;
2. Средний уровень. Источник информации относится к этой категории, если содержит большее количество критичных конфиденциальных сведений. Для персональных данных: от четырех прямых идентификаторов, исключая данные карт и паспортных данных. Для банковской тайны: от четырех прямых идентификаторов, исключая данные расчетных счетов и движения средств по ним. Для коммерческой тайны: описание алгоритмов, реализованных в закупленных или продаваемых программных и проектных продуктов;
3. Высокий уровень. Основной критерий, по которому исследуемый источник относится к данной категории – широкий набор высокочувствительных данных о субъекте информации. Для персональных данных: любое количество прямых идентификаторов, включая данные карт и паспортных данных. Для банковской тайны: любое количество прямых идентификаторов, включая данные расчетных счетов и движения средств по ним. Для коммерческой тайны: публикация части или полного программного кода, используемого в закупленных или продаваемых программных и проектных продуктов.

Последняя выделенная угроза от совершения инцидента разглашения конфиденциальной информации в открытых источниках – финансовые потери, связанные с нарушением законодательства. Для данного вида угроз также выделено 3 уровня уязвимости:

1. Низкий уровень. К данной категории относится информация, содержащая небольшое количество сведений о субъекте информации, но наиболее резонансное с точки зрения реакции субъектов информации. Для персональных данных: публикации содержат от одного до трех фактов. Для банковской тайны: от одного до трех характерных идентификаторов субъекта данных. Для коммерческой тайны: описание принципов работы продуктов, которые поставлены сторонней организацией и заключено соглашение о неразглашении информации;
2. Средний уровень. Источник информации относится к этой категории, если содержит большее количество резонансных конфиденциальных сведений. Для персональных данных: публикации содержат более четырех фактов. Для банковской тайны: публикации содержат более четырех фактов, характеризующих субъекта данных. Для коммерческой тайны: описание алгоритмов работы продуктов, которые поставлены сторонней организацией и заключено соглашение о неразглашении информации;
3. Высокий уровень. Основной критерий, по которому исследуемый источник относится к данной категории – широкий набор высокочувствительных данных о субъекте информации. Для персональных данных: любое количество прямых идентификаторов, но при этом обязательно наличие данных карт и паспортных данных. Для банковской тайны: любое количество прямых идентификаторов, обязательное наличие данные расчетных счетов и движения средств по ним. Для коммерческой тайны: публикация части или полного программного кода, используемого в продуктах, которые поставлены сторонней организацией и заключено соглашение о неразглашении информации.

Средствами искусственного интеллекта (ИИ) и машинного обучения (ML, machine learning) выявлены сущности, относящие публикацию (текст из открытого источника информации) к одному из трех видов конфиденциальной информации и одному из четырех типов событий угроз, с точки зрения категорирования чувствительных данных по заранее определенным критериям [17-22].

Следующая категориальная величина, используемая для оценки уровня риска от инцидентов разглашения конфиденциальной информации – это уровень критичности с точки зрения последствий от события – угрозы. Разделение по данным категориям производится на основе количества инцидентов, связанных с разглашением конфиденциальной информации каждого из трех типов (персональные данные, банковская тайна и коммерческая тайна) к общему количеству публикаций в рассматриваемом диапазоне времени (разделение по

временным диапазонам описано дальше). В таблице 2 представлена шкала вероятностей для реализации последствий от наступления событий-угроз и описание этих последствий. Предложенные значения процентов получены на основании процентильного распределения.

Таблица 2. **Шкала вероятностей для реализации последствий**

Table 2. **Probability scale for the realization of consequences**

Последствия	Обоснование	Вероятность реализации последствий
Можно пренебречь воздействием	События не ведут к потерям для организации либо ведут к минимальным потерям, которые не являются существенными	0,05
Последствия можно легко устранить	Случившиеся инциденты информационной безопасности, связанные с разглашением чувствительной информации, ведут к последствиям, которые можно легко устранить, запросив удаление конкретных публикаций, например, или заплатив минимальные штрафы.	0,1
Устранение требует умеренных трат	События – угрозы по количеству в одном источнике и относящиеся к определенному типу информации ведут к затратам на компенсацию последствий в рамках умеренных трат.	0,15
Серьезные последствия	Инциденты разглашения конфиденциальной информации появляются достаточно часто и ведут к серьезным последствиям с точки зрения их компенсации.	0,2
Критические последствия	Угрозы являются критическими и могут даже приостановить основную деятельность организации, например, в случае выявления такого объема инцидентов со стороны регуляторов.	0,5

В данном исследовании разделение на диапазоны времени обусловлено экспертным мнением и эмпирическими наблюдениями, получены следующие временные периоды:

1. Один месяц с момента публикации информации. Данный диапазон учитывает мгновенную реакцию на инциденты разглашения конфиденциальной информации в открытых источниках сети интернет. Реагирование субъектов данных и государственных регуляторов на инциденты, случившиеся за последний месяц наиболее вероятно, задается величиной равной 0,4;
2. Данный диапазон учитывает чуть более старые события: опубликованные в период от одного до восьми месяцев относительно текущего момента времени. Такие ограничения на диапазон обусловлены тем, что реакция на некритические инциденты не является мгновенной. Вероятность такой реакции задается значением равным 0,3, поскольку если реакция не поступила сразу, то возможно, последствия от такого разглашения не являются критическими;
3. Последний диапазон задается следующим временным интервалом: более восьми месяцев относительно текущего момента времени. Определение именно такого временного диапазона обусловлено наблюдением, массовый сбор информации, ее заимствование и применение для получения выгод требует некоторых временных затрат. По опыту экспертов информационной безопасности в Сбере, последствия наступают спустя восемь – девять месяцев. С другой стороны, если последствия не наступили сразу, то, возможно, они не наступят совсем[12]. Поэтому реакция на такие инциденты также имеет достаточно высокую вероятность для достаточно давних событий и задается величиной равной 0,3.

Данная таблица отражает взаимосвязь между временем публикации, относительно текущего момента времени и вероятности наступления последствий за случившиеся инциденты разглашения конфиденциальной информации.

В качестве параметров предложенной модели используются:

1. За величину отражающую вероятность угрозы с точки зрения влияния разглашения критической информации на продолжение основной деятельности организации выбрано распределение, отраженной в таблице 2 выше. Данные уровни заданы на основании отдаленности событий от текущего момента;
2. В качестве вероятностей уязвимостей с точки зрения возможности реализовать инцидент разглашения конфиденциальной информации в открытых источниках принято использовать распределение из пяти значений вероятности, обоснования полученных величин, описанных ранее, данное распределение вероятностей получено путем процентильного распределения и экспертных оценок;
3. За величину потерь принято использовать величины из таблицы 1, значение выбирается в зависимости от условий, под которые попадает набор инцидентов и категорирования информации.

[12] Коряковский Д., Теренин А. ИИ на страже банковских данных: опыт Сбербанка // BIS Journal – Информационная безопасность банков. 2020. № 2(37). С. 42-48. URL: https://ib-bank.ru/bisjournal/post/1319 (дата обращения: 07.09.2022); Гарбузов Г., Теренин А. ИИ на страже банковских данных – 2: опыт «Сбербанка» // BIS Journal – Информационная безопасность банков. 2020. № 4(39). URL: https://ib-bank.ru/bisjournal/post/1469 (дата обращения: 07.09.2022); Бабак Н. Г., Крюков А. Ф. Защита информации в операционной системе Android // Международный журнал информационных технологий и энергоэффективности. 2019. Т. 4, № 1(11). С. 21-26. URL: https://www.elibrary.ru/item.asp?id=37356369 (дата обращения: 07.09.2022); Бабак Н. Г. Способы распознавания и обезличивания персональных данных в автоматизированных системах // Радиоэлектроника, электротехника и энергетика : тезисы докладов Двадцать восьмой межд. НТК студентов и аспирантов. М. : РАДУГА, 2022. С. 182. URL: https://www.elibrary.ru/item.asp?id=48312089 (дата обращения: 07.09.2022).

В соответствии с исследуемым подходом к оценке риска для категориальных значений, получено, что допустимый уровень риска = 0,51504 в нормированной к 1 шкале.

Таблица 3. **Взаимосвязь вероятности последствий и давности публикации**

Table 3. **The relationship between the likelihood of consequences and the prescription of publication**

Временной диапазон / Вероятность последствий	1 месяц с текущего момента времени	От 1 до 8 месяцев с текущего момента времени	Более 8 месяцев с текущего момента времени
p	0,4	0,3	0,3

Полученные результаты

Для оценки текущих уровней риска для открытых источников от распространения конфиденциальной информации используется объективная вероятность (выраженная в относительной частоте появления какого-либо инцидента информационной безопасности к общему объему публикаций).

Для оценки прогнозных значений принято использовать субъективную вероятность (своего рода мера уверенности некоторого эксперта или группы экспертов в том, что данный инцидент в действительности будет реализован). Принятый способ представления таких вероятностей – распределение случайной величины.

Применив смоделированный подход к расчету уровня риска от разглашения конфиденциальной информации и используя формулы 1-4, полученные результаты представлены в таблице 4. Пример расчета для уровня риска из-за разглашения персональных данных в социальной сети ВКонтакте.

Таблица 4. **Полученные значения для социальной сети ВКонтакте для оценки уровня риска от разглашения персональных данных**

Table 4. **The obtained values for VKontakte social network for risk assessment of personal data disclosure**

Вид информации	Потери	Вероятность угрозы	0,05	0,1	0,15	0,2	0,5
Персональные данные (P=0.155)	Репутационные потери (P=0.3)	0,3	1	2	3	4	5
		0,3	2	3	4	5	6
		0,4	3	4	5	6	7
	Финансовые потери – штрафы (P=0.2)	0,3	1	2	3	4	5
		0,3	2	3	4	5	6
		0,4	3	4	5	6	7
	Финансовые потери – потеря клиентов или снижение конкурентоспособности (P=0.3)	0,3	1	2	3	4	5
		0,3	2	3	4	5	6
		0,4	3	4	5	6	7
	Нарушение законодательства (P=0.2)	0,3	1	2	3	4	5
		0,3	2	3	4	5	6
		0,4	3	4	5	6	7

Для того, чтобы определить уровень риска, необходимо применить формулы 1-4 со значениями из таблицы 4:

$$R_{ПД}^{vk} = 0{,}155 \times (0{,}3 \times (1 \times 0{,}05 + 2 \times 0{,}1 + 3 \times 0{,}15 + 4 \times 0{,}2 + 5 \times 0{,}5) + \\ + 0{,}3 \times (2 \times 0{,}05 + 3 \times 0{,}1 + 4 \times 0{,}15 + 5 \times 0{,}2 + 6 \times 0{,}5) + \\ + 0{,}4 \times (3 \times 0{,}05 + 4 \times 0{,}1 + 5 \times 0{,}15 + 6 \times 0{,}2 + 7 \times 0{,}5))$$

Исходя из полученных результатов сформирована шкала критичности по уровням риска, относительно допустимого уровня риска:

1. Низкий уровень (значение не превышает допустимый уровень) риска для источника с точки зрения разглашения конфиденциальной информации. Для таких источников можно пренебречь введением мер по предотвращению появления инцидентов информационной безопасности, связанных с публикацией чувствительной информации. Соответствует зеленому цвету на рисунке 1;
2. Средний уровень (значение превышает допустимый уровень риска, но меньше значения 0,7) риска для открытого источника информации, то есть внедрение мер по предотвращению инцидентов разглашения конфиденциальной информации является важным, но не критичным с точки зрения срочности внедрения этих мер. Соответствует оранжевому цвету на рисунке 1;
3. Высокий уровень (превышение значения 0,7) риска для исследуемого открытого источника информации. Внедрение мер по уменьшению рисков является не только важным, но и срочным. В первую очередь необходимо определить именно такие источники информации. На рисунке 1 такие источники выделены красным цветом.

Таблица 5. **Оцененные уровни риска**

Table 5. **Assessed risk levels**

Категорирование информации / Источники информации	Персональные данные	Банковская тайна	Коммерческая тайна
	Уровень риска		
ВКонтакте	0,79053	0,69673	0,21209
rbc.ru	0,49685	0,68136	0,17505
Banki.ru	0,59804	0,88450	0,26338

Полученные результаты отражают текущее положение уровней риска для рассматриваемых открытых источников. Для снижения полученных показателей риска рассмотрены следующие меры:

1. Повышение уровня грамотности;
2. Формализация требований для отнесения информации к определенному виду;
3. Внедрение сервиса по проверке и обезличиванию информации.

Данные меры могут быть скорректированы и адаптированы для различных типов информации, исходя из требований к ее хранению и обработке и деятельности самой организации, для которой производится анализ [23-29].

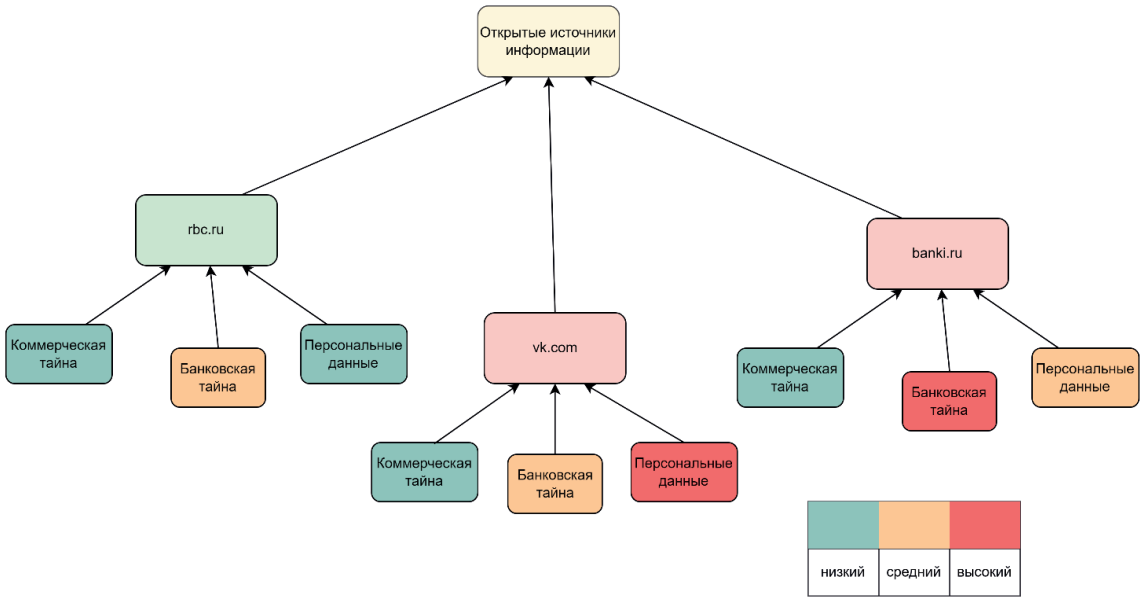


Рис. 1. Текущие уровни риска для рассматриваемых открытых источников и 3-х типов информации

Fig. 1. Current risk levels for considered open sources and 3 types of information

При условии применения рекомендуемых мер по сохранению конфиденциальности чувствительных данных, проведена еще одна итерация, по оценке количественных показателей уровня риска для открытых источников. Полученные прогнозные значения представлены в таблице 6 и рисунке 2.

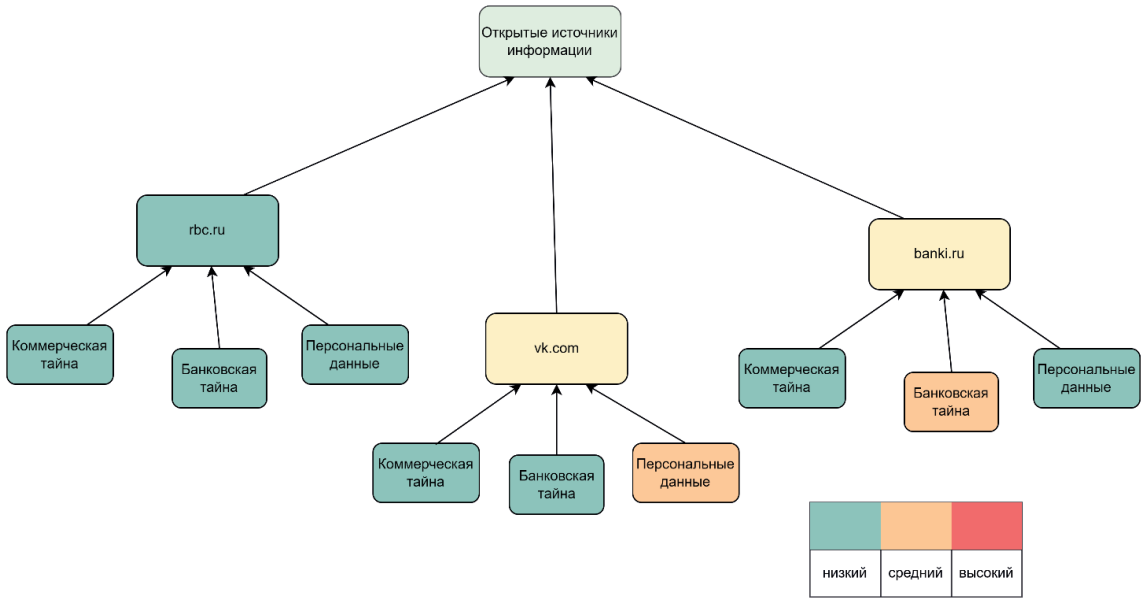
Таблица 6. **Прогнозные значения риска**

Table 6. **Forecast risk values**

Категорирование информации / Источники информации	Персональные данные	Банковская тайна	Коммерческая тайна
	Уровень риска		
ВКонтакте	0,61695	0,3846	0,1629
rbc.ru	0,29248	0,11222	0,14871
Banki.ru	0,50944	0,59685	0,21166

По полученным прогнозным значениям можно сделать вывод, что корректировка уровней риска от разглашения конфиденциальной информации для открытых источников информации возможна. Главное определить, что такая необходимость существует, разработать и адаптировать меры по сохранению конфиденциальности информации, которая не предназначена для публичной огласки в информационных ресурсах.

При сравнении рисунка 1 и рисунка 2, на которых представлены направленные графы, по уровням критичности для трех типов конфиденциальной информации в исследуемых открытых источниках наглядно показано, как уменьшается общий уровень риска для источников информации, при условии разработки и соблюдения мер по предотвращению разглашения конфиденциальной информации. Количественное подтверждение данных результатов отражено в таблице 5 и таблице 6.



Р и с. 2. Прогнозируемый уровень риска для анализируемых источников для 3-х типов информации

F i g. 2. Predicted risk level for analyzed sources for 3 types of information

Таким образом в данном исследовании оценены, по предложенному авторскому алгоритму, значения уровня риска от разглашения непубличной информации для трех источников информации и выделенных типов конфиденциальной информации. Исследование показывает, что существует возможность обнаружить «слабое место» с точки зрения разглашения конфиденциальной информации, и задуматься о развитии этого направления защиты информации. Важно отметить, что использование предложенного подхода носит рекомендательный характер и относится в первую очередь к крупным организациям, которые обладают большим количеством данных о клиентах и контрагентах, а также имеющим особые регуляторные требования.

Заключение

В результате проведенного анализа оценены уровни риска открытых источников, проведен сравнительный анализ между источниками по видам конфиденциальной информации, выявлены проблемные места (наиболее опасные с точки зрения разглашенной информации). Представлен алгоритм и математическая модель, которые позволяют провести эту оценку, а также, предложены рекомендации по усилению средств защиты конфиденциальной информации.

Для определения необходимости регулярного мониторинга систем (на примере открытых источников) на наличие неправомерно разглашенной конфиденциальной информации в ходе проведенного исследования были решены следующие задачи:

1. Сформированы требования к системе оценки рисков источников информации от разглашения разных видов конфиденциальной информации.
2. Разработана методика оценки риска разглашения конфиденциальной информации в системе, не предполагающей этого. Важно отметить, что данная методика должна использоваться с адаптацией под специфику деятельности организации, которая заинтересована в оценке риска исследуемого источника информации.
3. На рассматриваемом примере выявлено, что существует необходимость в анализе, в том числе, открытых источников.
4. Разработан подход к оценке рисков разглашения конфиденциальной информации в источниках информации. В использовании данного подхода возможна вариативность.
5. Получены и проинтерпретированы количественные оценки уровней риска, а также сформирована шкала, позволяющая оценить степень риска источников информации.

References

[1] Alshaikh M. Developing cybersecurity culture to influence employee behavior: A practice perspective. *Computers & Security.* 2020;98:102003. doi: https://doi.org/10.1016/j.cose.2020.102003

[2] Ameen N., Tarhini A., Shah M.H., Madichie N., Paul J., Choudrie J. Keeping customers' data secure: A cross-cultural study of cybersecurity compliance among the Gen-Mobile workforce. *Computers in Human Behavior*. 2021; 114:106531. doi: https://doi.org/10.1016/j.chb.2020.106531

[3] Paul J.A., Zhang M. Decision support model for cybersecurity risk planning: A two-stage stochastic programming framework featuring firms, government, and attacker. *European Journal of Operational Research*. 2021;291(1):349-64. doi: https://doi.org/10.1016/j.ejor.2020.09.013

[4] Singh J., Crisafulli B., Xue M.T. 'To trust or not to trust': The impact of social media influencers on the reputation of corporate brands in crisis. *Journal of Business Research*. 2020;119:464-80. doi: https://doi.org/10.1016/j.jbusres.2020.03.039

[5] Naarttijärvi M. Balancing data protection and privacy – The case of information security sensor systems. *Computer Law & Security Review*. 2018;34(5):1019-1038. doi: https://doi.org/10.1016/j.clsr.2018.04.006

[6] Tikkinen-Piri C., Rohunen A., Markkula J. EU General Data Protection Regulation: Changes and implications for personal data collecting companies. *Computer Law & Security Review*. 2018;34(1):134-153. doi: https://doi.org/10.1016/j.clsr.2017.05.015

[7] Steppe R. Online price discrimination and personal data: A General Data Protection Regulation perspective. *Computer Law & Security Review.* 2017;33(6):768-85. doi: https://doi.org/10.1016/j.clsr.2017.05.008

[8] Borgesius F.Z., Poort J. Online Price Discrimination and EU Data Privacy law. *Journal of Consumer Policy.* 2017;40(3):347-366. doi: https://doi.org/10.1007/s10603-017-9354-z

[9] Štitilis D., Laurinaitis M. Treatment of biometrically processed personal data: Problem of uniform practice under EU personal data protection law. *Computer Law & Security Review*. 2017;33(5):618-628. doi: https://doi.org/10.1016/j.clsr.2017.03.012

[10] Malatras A., Sanchez I., Beslay L., Coisel I., Vakalis I., D'Acquisto G., Sanchez M.G., Grall M., Hansen M., Zorkadis V. Pan-European personal data breaches: Mapping of current practices and recommendations to facilitate cooperation among Data Protection Authorities. *Computer Law & Security Review*. 2017;33(4):458-469. doi: https://doi.org/10.1016/j.clsr.2017.03.013

[11] Mantelero A. Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection. *Computer Law & Security Review*. 2016;32(2):238-55. doi: https://doi.org/10.1016/j.clsr.2016.01.014

[12] Li Y., Saxunová D. A perspective on categorizing personal and sensitive data and the analysis of practical protection regulations. *Procedia Computer Science*. 2020;170:1110-1115. doi: https://doi.org/10.1016/j.procs.2020.03.060

[13] Mousavi R., Chen R., Kim D.J., Chen K. Effectiveness of privacy assurance mechanisms in users' privacy protection on social networking sites from the perspective of protection motivation theory. *Decision Support Systems*. 2020;135:113323. doi: https://doi.org/10.1016/j.dss.2020.113323

[14] Zhao J., Yan Q., Li J., Shao M., He Z., Li B. TIMiner: Automatically extracting and analyzing categorized cyber threat intelligence from social data. *Computers & Security*. 2020;95:101867. doi: https://doi.org/10.1016/j.cose.2020.101867

[15] Choi J.P., Jeon D.S., Kim B.C. Privacy and personal data collection with information externalities. *Journal of Public Economics*. 2019;173:113-124. doi: https://doi.org/10.1016/j.jpubeco.2019.02.001

[16] Arooj A., Farooq M.S., Akram A. et al. Big Data Processing and Analysis in Internet of Vehicles: Architecture, Taxonomy, and Open Research Challenges. *Archives of Computational Methods in Engineering*. 2022;29(2):793-829. doi: https://doi.org/10.1007/s11831-021-09590-x

[17] Fisher J., Vlachos A. Merge and Label: A Novel Neural Network Architecture for Nested NER. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics; 2019. p. 5840-5850. doi: https://doi.org/10.18653/v1/P19-1585

[18] Mayhew S., Tsygankova T., Roth D. ner and pos when nothing is capitalized. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Hong Kong, China: Association for Computational Linguistics; 2019. p. 6256-6261. Available at: https://aclanthology.org/D19-1650.pdf (accessed 07.09.2022).

[19] Park J.-S., Kim G.-W., Lee D.-H. Sensitive Data Identification in Structured Data through GenNER Model based on Text Generation and NER. In: Proceedings of the 2020 International Conference on Computing, Networks and Internet of Things (CNIOT2020). New York, NY, USA: Association for Computing Machinery; 2020. p. 36-40. doi: https://doi.org/10.1145/3398329.3398335

[20] Hassan F., Domingo-Ferrer J., Soria-Comas J. Anonymization of Unstructured Data via Named-Entity Recognition. In: Torra V., Narukawa Y., Aguiló I., González-Hidalgo M. (eds.) Modeling Decisions for Artificial Intelligence. MDAI 2018. Lecture Notes in Computer Science. Vol. 11144. Cham: Springer; 2018. p. 296-305. doi: https://doi.org/10.1007/978-3-030-00202-2_24

[21] Guamán D.S., Ferrer X., del Alamo J.M., Such J. Automating the GDPR Compliance Assessment for Cross-border Personal Data Transfers in Android Applications. *arXiv*:2103.07297. 2021. doi: https://doi.org/10.48550/arXiv.2103.07297

[22] Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser Ł., Polosukhin I. Attention is All you Need. In: Guyon I., Von Luxburg U., Bengio S., Wallach H., Fergus R., Vishwanathan S., Garnett R. (eds.) Advances in Neural Information Processing Systems. vol. 30. Long Beach, CA, USA: Curran Associates, Inc.; 2017. Available at: https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html (accessed 07.09.2022).

[23] Branley-Bell D., Coventry L., Sillence E. Promoting Cybersecurity Culture Change in Healthcare. In: The 14th PErvasive Technologies Related to Assistive Environments Conference (PETRA 2021). New York, NY, USA: Association for Computing Machinery; 2021. p. 544-549. doi: https://doi.org/10.1145/3453892.3461622

[24] Mwim E.N., Mtsweni J. Systematic Review of Factors that Influence the Cybersecurity Culture. In: Clarke N., Furnell S. (eds.) Human Aspects of Information Security and Assurance. HAISA 2022. IFIP Advances in Information and Communication Technology. Vol. 658. Springer, Cham; 2022. p. 147-172. doi: https://doi.org/10.1007/978-3-031-12172-2_12

[25] Corradini I. Building a Cybersecurity Culture. In: Building a Cybersecurity Culture in Organizations. Studies in Systems, Decision and Control. Vol. 284. Cham: Springer; 2020. p. 63-86. doi: https://doi.org/10.1007/978-3-030-43999-6_4

[26] Uchendu B., Nurse J.R., Bada M., Furnell S. Developing a cyber security culture: Current practices and future needs. *Computers & Security*. 2021;109:102387. doi: https://doi.org/10.1016/j.cose.2021.102387

[27] Aiken G.M. Cybersecurity and productivity: has a cybersecurity culture gone too far? In: ASBBS Proceedings of the 26th Annual Conference. San Diego: American Society of Business and Behavioral Sciences; 2019. p. 13-23.

[28] Blum D. Executive Overview. In: Rational Cybersecurity for Business. Berkeley, CA: Apress; 2020. p. 1-29. doi: https://doi.org/10.1007/978-1-4842-5952-8_1

[29] Babak N.G., Kryukov A.F. Mobile Application for Visualization of the Advertising Booklet Using Augmented Reality. In: 2018 IV International Conference on Information Technologies in Engineering Education (Inforino). Moscow, Russia: IEEE Computer Society; 2018. p. 1-4. doi: https://doi.org/10.1109/INFORINO.2018.8581841

Поступила 07.09.2022; одобрена после рецензирования 05.10.2022; принята к публикации 14.10.2022.
Submitted 07.09.2022; approved after reviewing 05.10.2022; accepted for publication 14.10.2022.

Об авторах:

Шаброва Анастасия Игоревна, архитектор по защите данных, Департамент кибербезопасности, ПАО «Сбербанк России» (117312, Российская Федерация, г. Москва, ул. Вавилова, д. 19), **ORCID: https://orcid.org/0000-0002-4315-3061,** AIgShabrova@sber.ru

Теренин Алексей Алексеевич, управляющий директор, Департамент кибербезопасности, ПАО «Сбербанк России» (117312, Российская Федерация, г. Москва, ул. Вавилова, д. 19), кандидат технических наук, **ORCID: https://orcid.org/0000-0002-6242-6117,** Terenin.A.Alek@sberbank.ru

Бабак Никита Григорьевич, главный эксперт по защите данных, Департамент кибербезопасности, ПАО «Сбербанк России» (117312, Российская Федерация, г. Москва, ул. Вавилова, д. 19), **ORCID: https://orcid.org/0000-0001-7129-1018,** NGBabak@sber.ru

Все авторы прочитали и одобрили окончательный вариант рукописи.

About the authors:

Anastasiia I. Shabrova, Data Protection Architect, Cybersecurity Department, PJSC "Sberbank of Russia" (19 Vavilov St., Moscow 117312, Russian Federation), **ORCID: https://orcid.org/0000-0002-4315-3061,** AIgShabrova@sber.ru

Aleksey A. Terenin, Managing Director, Cybersecurity Department, PJSC "Sberbank of Russia" (19 Vavilov St., Moscow 117312, Russian Federation), Cand.Sci. (Eng.), **ORCID: https://orcid.org/0000-0002-6242-6117,** Terenin.A.Alek@sberbank.ru

Nikita G. Babak, Chief Data Protection Expert, Cybersecurity Department, PJSC "Sberbank of Russia" (19 Vavilov St., Moscow 117312, Russian Federation), **ORCID: https://orcid.org/0000-0001-7129-1018,** NGBabak@sber.ru

All authors have read and approved the final manuscript.