СИСТЕМА ИЗМЕРЕНИЯ ЗАЩИЩЕННОСТИ ИНФОРМАЦИИ И ПЕРСОНАЛЬНЫХ ДАННЫХ ДЛЯ УСТРОЙСТВ ИНТЕРНЕТА ВЕЩЕЙ

Федорченко Е.В.[1], Новикова Е.С.[2], Котенко И.В.[3], Гайфулина Д.А.[4], Тушканова О.Н.[5], Левшун Д.С.[6], Мелешко А.В.[7], Муренин И.Н.[8], Коломеец М.В.[9]

Цель статьи: *устранение противоречия, состоящего в существующей потребности в наборе простых и понятных показателей защищенности информации и персональных данных для пользователей устройств интернета вещей и их производителей, и отсутствием такого набора, объединяющего взаимосвязанные показатели защищенности информации и персональных данных, а также алгоритмы их вычисления и формирования понятной и объективной интегральной оценки, за счет разработки системы измерения защищенности информации и персональных данных для устройств интернета вещей.*

Метод исследования: *теоретический и системный анализ для определения и классификации показателей защищенности информации и персональных данных, семантический анализ для построения семантической модели сценариев обработки персональных данных, методы аналитического моделирования для формирования последовательностей атакующих действий, методы анализа журналов событий, статистические методы и методы машинного обучения для выявления аномалий в поведении устройств, разработка баз данных и программного кода для реализации системы измерения защищенности.*

Полученный результат: *предложена система измерения защищенности информации и персональных данных для пользователей устройств интернета вещей и их производителей. Используя доступные данные об устройствах интернета вещей, предложенная система позволяет автоматически вычислять показатели защищенности информации и персональных данных и формировать на их основе интегральную оценку защищенности информации и персональных данных. В рамках данной системы разработана иерархия показателей защищенности информации и персональных данных для устройств интернета вещей. Предложенные показатели вычисляются на основе статической и динамической информации об устройстве и его поведении. Разработаны оригинальные алгоритмы вычисления предложенных показателей, в том числе интегрального показателя защищенности информации и персональных данных. Определена архитектура системы измерения защищенности информации и персональных данных, объединяющая компоненты, реализующие предложенные алгоритмы вычисления показателей. Функционирование системы продемонстрировано на примерах.*

Область применения предложенной системы – разработанная система измерения защищенности может применяться производителями устройств и систем интернета вещей для анализа уровня их защищенности, а также для предоставления пользователям простых и понятных показателей защищенности информации и персональных данных.

1 Федорченко (Дойникова) Елена Владимировна, кандидат технических наук, старший научный сотрудник, ФГБУН «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН), г. Санкт-Петербург, Россия. E-mail: doynikova@comsec.spb.ru

2 Новикова Евгения Сергеевна, кандидат технических наук, старший научный сотрудник, ФГБУН «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН), г. Санкт-Петербург, Россия. E-mail: novikova@comsec.spb.ru

3 Котенко Игорь Витальевич, доктор технических наук, профессор, главный научный сотрудник и руководитель лаборатории проблем компьютерной безопасности, ФГБУН «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН), г. Санкт-Петербург, Россия. E-mail: ivkote@comsec.spb.ru

4 Гайфулина Диана Альбертовна, аспирант, младший научный сотрудник, ФГБУН «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН), г. Санкт-Петербург, Россия. E-mail: gaifulina@comsec.spb.ru

5 Тушканова Ольга Николаевна, кандидат технических наук, старший научный сотрудник, ФГБУН «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН), г. Санкт-Петербург, Россия. E-mail: tushkanova@comsec.spb.ru

6 Левшун Дмитрий Сергеевич, кандидат технических наук, старший научный сотрудник, ФГБУН «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН), г. Санкт-Петербург, Россия. E-mail: levshun@comsec.spb.ru

7 Мелешко Алексей Викторович, аспирант, младший научный сотрудник, ФГБУН «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН), г. Санкт-Петербург, Россия. E-mail: meleshko.a@iias.spb.su

8 Муренин Иван Николаевич, аспирант, младший научный сотрудник, ФГБУН «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН), г. Санкт-Петербург, Россия. E-mail: imurenin@gmail.com

9 Коломеец Максим Вадимович, Ph.D. в компьютерных науках, старший научный сотрудник, ФГБУН «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН), г. Санкт-Петербург, Россия. E-mail: kolomeec@comsec.spb.ru

Научная новизна: *разработана иерархия статических и динамических показателей защищенности информации и персональных данных для устройств интернета вещей; предложен подход к оцениванию защищенности устройств интернета вещей на основе выделенных показателей и доступных данных безопасности; разработаны оригинальные алгоритмы вычисления отдельных показателей; разработаны алгоритмы вычисления интегральных оценок с учетом доступных данных безопасности.*

Вклад: *Федорченко Е.В. – разработка подхода, иерархии показателей и архитектуры системы, постановка задач по компонентам и участие в их разработке, Новикова Е.С. – компонент вычисления оценки рисков, связанных с обработкой персональных данных, компонент вычисления интегральных оценок рисков, Котенко И.В. – общее руководство, постановка задачи, архитектура системы, Гайфулина Д.А. – компонент обработки и интеграции журналов событий, Тушканова О.Н., Муренин И.Н. – компонент вычисления динамической оценки на основе статистических методов и методов машинного обучения, Левшун Д.С. – база данных показателей, компонент вычисления статической оценки, Мелешко А.В. – компонент оценки удобочитаемости, Коломеец М.В – компонент вычисления оценки рисков, связанных с обработкой персональных данных, на основе файлов *.apk, компонент вычисления динамической оценки с учетом возможных последовательностей атак. Все авторы участвовали в написании статьи.*

Ключевые слова: *оценивание защищенности, показатели защищенности информации и персональных данных, интегральные оценки, статическая информация, динамическая информация, семантика, онтология, кибератака, информационная система, интеллектуальный анализ данных, аномалии, цепочки атак.*

DOI:10.681/2311-3456-2022-5-28-46

1. Введение

В настоящее время системы и устройства интернета вещей используются повсеместно, а количество кибератак на такие устройства растет. Вследствие этого как для производителей устройств, так и для их пользователей, становится актуальным вопрос, насколько защищены используемые ими устройства от кибератак. Ответить на этот вопрос могли бы простые и объективные показатели защищенности информации и персональных данных (ПДн).

Существуют контрольные списки требований защищенности устройств интернета вещей и показателей, указывающих на выполнение соответствующих требований. Например, система соответствия уровня защищенности для интернета вещей “IoT Security Compliance Framework 2.0” от организации IoT Security Foundation[10]. Кроме того, исследователями предложены различные подходы и методики вычисления отдельных показателей защищенности информации и персональных данных, таких как конфиденциальность, целостность и доступность информации (Confidentiality, Integrity, Availability - CIA)[11] [1], подлинность информации [2], защищенность персональных данных [3, 4, 5], прозрачность, удобочитаемость [6], надежность [7] и др.

Однако в настоящий момент не существует системы измерения защищенности для устройств интернета вещей, объединяющей взаимосвязанные показатели защищенности информации и персональных данных, а также алгоритмы их вычисления на основе «сырых» данных, и позволяющей сформировать на их основе понятную и объективную интегральную оценку. Таким образом, существует противоречие между потребностью в наборе простых и понятных показателей защищенности информации и персональных данных для пользователей устройств интернета вещей и их производителей, и существующими решениями, реализующими ограниченный набор измерений защищенности.

В данной статье предложена система измерения защищенности информации и персональных данных для пользователей устройств интернета вещей и их производителей, которая позволяет устранить обозначенное противоречие. Предложенная система дает возможность автоматически вычислять показатели защищенности информации и персональных данных на основе доступной информации об устройствах интернета вещей и формировать на их основе интегральную оценку защищенности информации и персональных данных. В рамках данной системы разработана иерархия показателей защищенности информации и персональных данных для устройств интернета вещей. Предложенные показатели вычисляются на

10 https://www.iotsecurityfoundation.org/best-practice-guidelines

11 https://www.first.org/cvss/specification-document

основе статической и динамической информации об устройстве и его поведении.

Выделены следующие показатели, вычисляемые на основе статической информации:

— статическая оценка рисков нарушения CIA (вычисляется на основе собственной критичности устройства с точки зрения конфиденциальности, целостности и доступности хранимой информации, возможности реализации кибератаки, и потенциального ущерба для конфиденциальности, целостности и доступности хранимой устройством информации с учетом ее критичности);

— оценка риска нарушения конфиденциальности ПДн, определяемая на основе политик безопасности персональных данных (вычисляется на основе удобочитаемости текста политик и онтологии политик);

— оценка риска нарушения конфиденциальности ПДн на основе APK (Android Package) (вычисляется с использованием описания и прав доступа APK), а также интегральная статическая оценка риска нарушения защищенности.

Выделены следующие показатели, вычисляемые на основе динамической информации:

— динамическая оценка риска нарушения защищенности персональных данных (вычисляется с учетом оценки, вычисленной на основе APK, и динамической информации об аномалиях, связанных с персональными данными, обнаруженными в журналах событий устройств и системы);

— динамическая оценка риска нарушения CIA (определяется на основе динамической оценки возможности реализации кибератак и статической оценки риска нарушения CIA);

— интегральная динамическая оценка риска нарушения защищенности.

Разработаны оригинальные алгоритмы вычисления предложенных показателей, в том числе алгоритм вычисления статической оценки риска нарушения защищенности CIA (оценки CIA), оценки риска нарушения защищенности персональных данных на основе онтологии, оценки на основе APK, динамической оценки риска нарушения защищенности персональных данных и CIA и интегрального показателя риска нарушения защищенности информации и персональных данных.

Определена архитектура системы измерения защищенности информации и персональных данных, объединяющая компоненты, реализующие предложенные алгоритмы вычисления показателей.

Работа системы продемонстрирована на примерах.

Статья организована следующим образом. В разделе 2 приведен анализ релевантных работ. В разделе 3 описывается предложенная система измерения защищенности информации и персональных данных, ее компоненты, иерархия показателей и алгоритмы их вычисления. В разделе 4 рассматривается реализация предложенной системы измерения защищенности, и приводится пример ее работы.

2. Релевантные работы

Измерению защищенности интернета вещей посвящено большое количество исследований. Во-первых, существуют рекомендации, определяющие основные принципы защищенности устройств интернета вещей, которые должны быть удовлетворены. Например, следует выделить систему оценивания кибер-защищенности (Cyber Assessment Framework, CAF)[12] и систему определения соответствия уровня защищенности интернета вещей (IoT Security Compliance framework)[13]. Такие рекомендации определяют, какие требования защищенности должны быть удовлетворены и измерены, но не определяют, каким образом.

На основе этих документов и стандартов безопасности авторы данной статьи выделили основные показатели защищенности устройств интернета вещей. К ним относится риск нарушения конфиденциальности, целостности и доступности информации (оценка CIA), риск нарушения подлинности информации, риск нарушения защищенности персональных данных, прозрачность информации и другие.

Исследователями были предложены различные подходы вычисления данных показателей. Например, конфиденциальность, целостность и доступность информации, хранимой или передаваемой устройством, может вычисляться на основе доступной информации об уязвимостях устройства. Известные уязвимости и их оценки по системе оценивания уязвимостей (Common Vulnerability Scoring System, CVSS)[14], отражающие вероятность использования уязвимости и ущерб для конфиденциальности, целостности и доступности информации, хранимой или передаваемой устройством, в случае успешного использования уязвимости, можно найти в открытых базах данных, таких как Национальная база уязвимостей (National

12 https://www.ncsc.gov.uk/collection/caf/cyber-assessment-framework

13 https://www.iotsecurityfoundation.org/wp-content/uploads/2018/12/IoTSF-IoT-Security-Compliance-Framework-Release-2.0-December-2018.pdf

14 https://www.first.org/cvss/specification-document

Vulnerability Database, NVD). Эти оценки можно использовать по отдельности или в рамках более сложных подходов, учитывающих связи между устройствами и уязвимостями [1]. Для измерения риска нарушения конфиденциальности информации и конфиденциальности персональных данных может использоваться подход, основанный на анализе прав доступа, предоставленных программному или аппаратному обеспечению устройств. Также существуют подходы к вычислению риска нарушения конфиденциальности персональных данных и прозрачности политик безопасности персональных данных (политик конфиденциальности), основанные на анализе текста политик. К ним относятся подходы, основанные на правилах [3], онтологии [4, 8, 9, 10, 11] и машинном обучении [5]. Для вычисления удобочитаемости применяются различные алгоритмы оценивания сложности текста [6]. Для вычисления оценки CIA и риска нарушения подлинности информации также могут использоваться подходы на основе машинного обучения [2].

Хотя существует большое количество исследований в области оценивания защищенности информации и конфиденциальности персональных данных устройств интернета вещей, на текущий момент не разработано всеобъемлющей системы, объединяющей различные показатели защищенности информации и конфиденциальности персональных данных устройств, алгоритмы их вычисления, а также позволяющей сформировать на основе данных показателей понятную и объективную интегральную оценку.

3. Система измерения защищенности информации и конфиденциальности персональных данных

Предлагаемая авторами статьи система измерения защищенности предназначена для сравнения устройств Интернета вещей с точки зрения защищенности с использованием набора выделенных показателей. В зависимости от типа входных данных предлагаемый набор показателей можно разделить на статические и динамические. Иерархия предлагаемых статических показателей представлена на рис. 1, а иерархия динамических показателей - на рис. 2.

Для вычисления показателей предлагается использовать следующие входные данные:

- программное обеспечение, установленное на устройстве;
- описание программного обеспечения;
- соответствующие политики конфиденциальности;
- названия приложений (файлы *.apk), установленных на устройствах;
- данные NVD о программном и аппаратном обеспечении устройства, такие как общее перечисление платформ (Common Platform Enumeration, CPE) и известные уязвимости (Common Vulnerabilities and Exposures, CVE);
- журнал событий (лог) нормального поведения устройств;
- спецификации устройств.

Для вычисления динамических показателей, отличающихся тем, что их значения изменяются во времени, помимо потоковых данных журналов событий

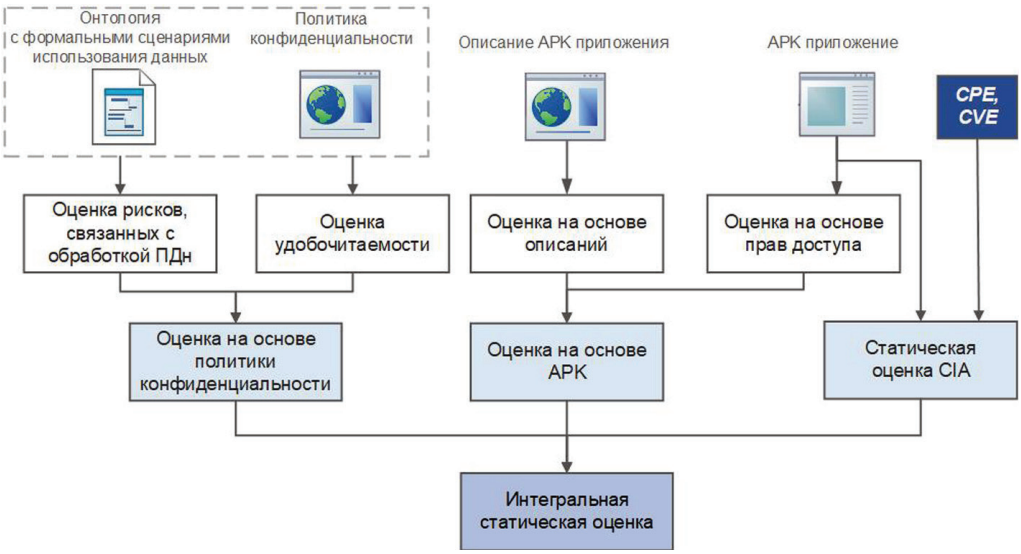

Рис. 1. Иерархия статических показателей защищенности информации и конфиденциальности персональных данных

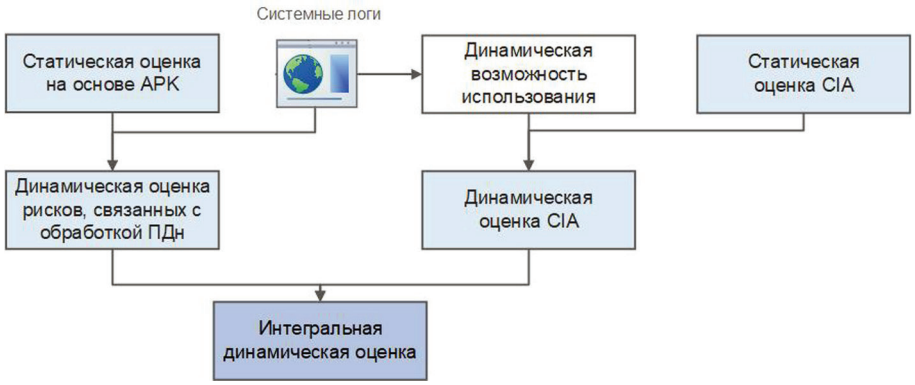

Рис. 2. Иерархия динамических показателей защищенности информации и конфиденциальности персональных данных

Рис. 3. Общая схема функционирования предлагаемой системы измерения защищенности

устройства предлагается использовать статические показатели (рис. 2).

На рис. 3 представлена общая схема предлагаемой системы.

Для вычисления выделенных показателей защищенности (оценок) разработаны алгоритмы, реализованные в рамках следующих компонентов системы измерения защищенности:

1) *компонент обработки и интеграции журналов,* реализующий анализ и объединение различных журналов событий;

2) *компонент вычисления статической оценки CIA* на основе оценок CVSS известных уязвимостей устройств;

3) *компонент вычисления оценки рисков, связанных с обработкой персональных данных, на основе APK* (файлов *.apk), характеризующей защищенность устройства с учетом запрошенных и необходимых прав доступа установленных приложений;

4) *компонент вычисления оценки рисков, связанных с обработкой персональных данных,* характеризующей риски конфиденциальности информации с учетом политики конфиденциальности устройства;

5) *компонент оценки удобочитаемости* политики конфиденциальности устройства для обеспечения всесторонней оценки рисков конфиденциальности, связанных с устройством;

6) *компонент вычисления динамической оценки CIA с учетом возможных последовательностей атак*, определяемых на основе конфигурации компьютерной и уязвимостей, которые могут быть использованы для компрометации устройства;

7) *компонент вычисления динамических оценок CIA и рисков, связанных с обработкой персональных данных,* на основе статистических методов и методов машинного обучения, используемых для обнаружения аномалий в поведении устройства;

8) *компонент вычисления интегральных оценок рисков нарушения защищенности и конфиденциальности*, включая интегральную оценку на основе политики конфиденциальности;

9) *база данных показателей.*

Алгоритмы вычисления показателей, реализуемых отдельными компонентами, подробно описаны в следующих подразделах.

3.1. Компонент вычисления статической оценки CIA

Вычисление статической оценки CIA основано на анализе известных уязвимостей программного и аппаратного обеспечения устройства и включает следующие основные этапы:

1) поиск уязвимостей устройства;

2) получение модифицированной оценки CVSS для найденных уязвимостей;

3) вычисление оценки CIA как максимальной оценки CVSS по всем обнаруженным уязвимостям устройства.

Информацию об уязвимостях устройств можно найти в открытых базах данных, таких как NVD, либо путем тестирования на проникновение. NVD содержит информацию об уязвимостях в формате CVE, связанном с форматом CPE, который представляет собой формальное описание программного и аппаратного обеспечения. Для поиска уязвимостей устройств в NVD необходимо сформировать список аппаратных средств, приложений или операционных систем в формате CPE [12].

Каждая запись CVE имеет оценку CVSS, отражающую основные характеристики уязвимостей. Эта оценка включает информацию об ущербе для CIA в случае успешного использования уязвимости и простоте использования уязвимости. В компоненте вычисления статической оценки CIA используется модифицированная оценка CVSS, учитывающая критичность анализируемого устройства.

3.2. Компонент вычисления оценки рисков, связанных с обработкой персональных данных, на основе APK

Компонент вычисления оценки рисков, связанных с обработкой персональных данных, на основе APK, реализует вычисление оценки рисков для Android-устройства с учетом запрошенных и необходимых прав доступа установленных на устройстве приложений. Описание приложения, доступное в магазине мобильных приложений (Google Play Store, Huawei App Gallery и т.д.), используется для прогнозирования набора прав доступа, которые требуются приложению для выполнения заявленных функций. Например, ожидаемо, что доступ к списку контактов пользователя потребуется приложению социальной сети, но подозрительно, если такой доступ требуется приложению-фонарику. Разница между прогнозируемыми и фактическими правами доступа приложений, установленных на устройстве, может служить основой для расчета рисков, связанных с обработкой персональных данных.

Выделим следующие виды оценок рисков, вычисляемых на основе прав доступа:

- оценка рисков, связанных с обработкой персональных данных, на основе описания APK, принимающая значения в диапазоне [0, 1.0];
- оценка рисков, связанных с обработкой персональных данных, на основе прав доступа APK, принимающая значения в диапазоне [0, 1.0];
- интегральная оценка, вычисляемая на основе APK, принимающая значения в диапазоне [0, 10].

Права доступа можно сгруппировать по типу (критичности) данных, с которым они могут быть связаны. Авторы данной статьи предлагают выделить восемь групп. Каждой группе присваивается вес $w \in \{0, 1, 2, 3, 4\}$, отражающий критичность прав доступа в контексте типов персональных данных, выделенных в GDPR[15]. Например, в группу прав доступа «Здоровье» входят права доступа к сенсорам (BODY_SENSORS). Обрабатываемая ими информация относится к категории персональных данных GDPR «Специальная». Поскольку это достаточно критичная информация, группе прав доступа «Здоровье» назначается высокий вес $w = 4$.

Формируется 2 двоичных вектора из 8 элементов каждый: первый вектор прав доступа формируется на основе описания приложения APK, а второй – на основе реальных прав доступа приложения APK. Каждый элемент вектора соответствует группе прав доступа. Для первого вектора элемент принимает значение 1, если на основе описания приложения сделан вывод, что соответствующие права доступа требуются для его функционирования. В противном случае элемент принимает значение 0. В случае второго вектора элемент принимает значение 1, если приложение имеет соответствующие права доступа. В противном случае элемент принимает значение 0.

Оценка рисков *PS*, связанных с обработкой персональных данных, на основе прав доступа APK, вычисляется с использованием вектора реальных прав доступа по следующей формуле:

15 https://gdpr-info.eu/

$$PS = \frac{\sum_{i=1}^{8}(A_i \times W_i)}{\sum_{i=1}^{8} W_i},$$

где A – вектор прав доступа;
W – вектор весов соответствующих прав доступа.

Оценка рисков *DS*, связанных с обработкой персональных данных, на основе описания АРК, вычисляется с использованием вектора отличий D между вектором прав доступа на основе описаний и вектором реальных прав доступа. Элемент вектора отличий принимает значение 1, если реальные права доступа не были спрогнозированы, а в остальных случаях – значение 0. Оценка рисков *DS* вычисляется по следующей формуле:

$$DS = \frac{\sum_{i=1}^{8}(D_i \times W_i)}{\sum_{i=1}^{8} W_i}.$$

Например, для вектора отличий $D = (0,0,1,0,0,0,0,0)$ и соответствующего вектора весов прав доступа $W = (4,2,2,2,2,2,1,0)$ оценка риска на основе описания *DS* вычисляется следующим образом:

$$DS = \frac{0*4+0*2+1*2+0*2+0*2+0*2+0*1+0*0}{4+2+2+2+2+2+1+0} = 0.13.$$

Интегральная оценка на основе АРК рассчитывается с использованием Алгоритма 1.

Алгоритм 1

```
1:  log_base = round(10 * PS)
2:  if log_base < e then
3:      ipps = PS * ln(1 + ln(1 + DS))
4:  else
5:      ipps = PS * log log_base (1 + ln(1 + DS))
6:  end if
7:  if ipps > 1 then
8:      ipps = 1
9:  end if
10: return ipps * 10
```

В основе алгоритма лежит оценка рисков *PS*, вычисленная на основе реальных прав доступа АРК. В рамках алгоритма вычисления интегральной оценки *PS* необходимо увеличить, если реальные права доступа отличаются от спрогнозированных на основе описания АРК (оценка риска *DS*). При этом интегральная оценка должна принимать значения от 0 до 10. При анализе значений интегральной оценки было выявлено, что в данном случае перемножение и сложение *PS* и *DS*, нормированных от 0 до 10, не подходит, так как не позволяет получить интегральную оценку, значения которой будут равномерно расти от 0 до 10, сохраняя распределение значений *PS*. Для того, чтобы удовлетворить данное требование, было принято решение использовать логарифмическую функцию. В алгоритме используется разное основание логарифма, так как когда *PS* превышает e/10, функция начинает слишком быстро расти. Использование основания log_base, динамически зависящего от параметра *PS*, позволяет замедлить этот рост. Использование этого основания в случае если log_base меньше или равно e ведет к отрицательному значению функции.

В представленном алгоритме *PS* принимает значения в диапазоне [0.0; 1.0], *DS* в диапазоне [0.0; 1.0], *ipps* – выходная интегральная оценка риска на основе прав доступа – в диапазоне [0.0; 10], *e* – Euler's number.

Оценка риска на основе прав доступа *PS* отражает риски, связанные с правами доступа, запрошенными приложением. Оценка риска на основе описания приложения *DS* отражает риски, которые рассчитываются на основе соответствия между правами доступа, запрошенным приложением, и его описанием.

Поэтому ее можно рассматривать как соответствие между запрашиваемыми данными и целями, для которых они собираются. Таким образом, если цели не ясны, и соответствующая оценка имеет высокое значение, необходимо увеличить интегральную оценку риска. Однако, если цели ясны, интегральная оценка определяется оценкой риска на основе прав доступа, поскольку риски, связанные с использованием персональных данных, все еще присутствуют.

3.3. Компонент оценки рисков, связанных с обработкой персональных данных

Данный компонент реализует методику оценки защищенности персональных данных, представленную в работе [13]. В ее основе лежит онтология, представляющая собой формализованное описание трех основных сценариев обработки персональных данных (сценариев использования): сбор и использование данных первыми лицами, распространение и передача данных третьим лицам и хранение данных.

Каждый сценарий использования персональных данных представлен множеством связанных между собой классов и свойств, которые определяют различные атрибуты заданного сценария использования. Например, они характеризуют тип собираемых или передаваемых персональных данных, цель их обработки, время хранения и т.д.

Основными понятиями (классами) онтологии являются класс «Данные» (Data) и его подклассы, такие как данные об учетной записи пользователя (User Account Data), данные о геолокации (Tracking Data), данные о приложении и устройстве (App&DevInfo), финансовые данные (User Financial Data) и т.д. Они определяют базовую оценку риска. Другие классы сценария использования персональных данных влияют на базовую оценку риска, увеличивая или снижая его. Таким образом, обобщенная формула вычисления оценки рисков, связанных с обработкой персональных данных в рамках отдельного сценария обработки данных, определяется следующим образом:

$$PDDataUsageScenarioRisk = PDRiskScoreBase * riskCoeff,$$

где *PDDataUsageScenarioRisk* – оценка риска, связанного с обработкой персональных данных, для конкретного сценария их использования, например хранения данных;

PDRiskScoreBase – базовая оценка риска, рассчитанная на основе типов персональных данных, используемых в сценарии, и их критичности;

riskCoeff – это коэффициент поправки риска, который определяется на основе экземпляров других классов сценария, т.е. цели использования, правовой основы их обработки, возможности отказаться от обработки.

Для расчета коэффициента поправки риска *riskCoeff* требуется выявить связанные классы, относящиеся к данному сценарию использования, исключая класс «Данные» и его подклассы. Каждый класс имеет подклассы, например класс «Время хранения» (Retention Time) имеет 4 подкласса: «Не определено» (Not Defined), «Определено» (Stated), «Бесконечно» (Indefinitely) и «Другое» (Other). Для каждого класса экспертным путем назначается уровень критичности по аналогии с уровнем критичности, определяемым для подклассов класса «Данные». Затем коэффициент *riskCoeff* вычисляется на основе значений уровня критичности каждого класса, выявленного для заданного сценария [13].

В текущей версии алгоритма оценки рисков, связанных с обработкой персональных данных, итоговая оценка на основе онтологии рассчитывается как среднее от оценок риска, определенных для каждого сценария использования, обнаруженного в политике конфиденциальности.

3.4. Компонент оценки удобочитаемости политик конфиденциальности

Оценка удобочитаемости (readability) относится к группе оценок, связанных с обработкой персональных данных, и отражает уровень простоты или сложности чтения какого-либо текста и, как следствие, сложность его понимания. В случае политик конфиденциальности сложность прочтения и понимания документа непосредственно связана с риском неправомочного использования персональных данных, так как пользователь дает согласие на их обработку без четкого понимания того, какие данные обрабатываются, какие цели их обработки и т.д. По этой причине риск удобочитаемости часто включается в оценки рисков, связанных с обработкой персональных данных.

В данной работе для расчета оценки удобочитаемости используется известная формула проверки сложности письменных текстов Flesch-Kincaid Grade Level readability (FKGLR).

Вывод об удобочитаемости конкретного текста делается на основании вычисленного значения FKGLR, исходя из следующих интервалов:

- FKGLR $\in [0, 6]$ – низкий уровень риска удобочитаемости, текст очень легко читать;
- FKGLR $\in (6, 10]$ – низкий уровень риска удобочитаемости, текст прост для среднестатистического читателя;
- FKGLR $\in (10, 12]$ – средний уровень риска удобочитаемости, текст несколько более сложен для среднестатистического читателя;
- FKGLR $\in (12, \infty]$ – высокий уровень риска удобочитаемости, текст сложный для опытного читателя, готового читать научные тексты.

3.5. Компонент вычисления динамической оценки CIA с учетом возможных последовательностей атакующих действий

Данный компонент позволяет учитывать не только отдельные атакующие действия при оценке защищенности устройства, но и последовательности атакующих действий. Каждое атакующее действие представляет собой использование уязвимости устройства. Последовательность атакующих действий формируется с учетом возможности последовательной эксплуатации уязвимостей устройства, а также уязвимостей, связанных с анализируемым устройством. Наличие потенци-

альной последовательности атакующих действий может увеличить риск нарушения CIA для устройства.

Например, уязвимость $v1$ устройства представляет низкий уровень риска, поскольку для ее использования требуются права пользователя. Но при наличии другой уязвимости $v2$, позволяющей получить необходимые права, можно сформировать последовательность атакующих действий $v1 \rightarrow v2$, которая представляет более высокий риск.

Для определения возможных последовательностей атакующих действий необходимо проанализировать конфигурацию компьютерной сети и состав программного и аппаратного обеспечения устройств. Компонент вычисления динамической оценки CIA с учетом возможных последовательностей атакующих действий использует в качестве входных данных журналы событий устройства и выполняет следующие шаги:

1. Выделение устройств, подключенных к одному и тому же шлюзу (определение подсетей).
2. Определение устройств в пределах каждой подсети (их программного и аппаратного обеспечения в формате CPE).
3. Нахождение уязвимостей устройств в формате CVE.
4. Определение пред- и постусловий использования уязвимостей на основе следующих показателей CVSS версии 3 (CVSSv3): вектор атаки (ВА), требуемые привилегии (права доступа) и полученные привилегии. Выделяется 5 групп уязвимостей на основе значений показателей CVSSv3. Они представлены в табл. 1.
5. Формирование последовательностей атакующих действий на основе выделенных групп уязвимостей и отношений между ними. На рис.4 представлены отношения между уязвимостями разных групп для одного хоста и разных хостов, определенные с учетом пред- и постусловий использования уязвимостей (табл.1). Группы выделены прямоугольниками и подписаны (V0-V4). Внутри прямоугольника перечислены пред- и постусловия использования уязвимостей соответствующей группы. Символ «!» обозначает операцию «НЕ».
6. Вычисление оценки CIA с учетом возможных последовательностей атакующих действий на устройство. Оценка CIA рассчитывается на основе оценок CVSSv3, как и в статическом случае. Ущерб для конфиденциальности, целостности и доступности в случае успешного выполнения последовательности атакующих действий определяется как максимальный ущерб в случае успешного выполнения отдельных атакующих действий (ущерб в случае успешного использования уязвимости согласно CVSSv3). Возможность успешного выполнения последовательности атакующих действий рассчитывается как произведение максимальной оценки ВА для уязвимостей последовательности, сложностей доступа для всех уязвимостей в последовательности и оценок требуемых привилегий и взаимодействия с пользователем согласно CVSSv3. Оценка CIA для последовательности рассчитывается на основе полученных оценок ущерба и возможности использования. Оценка CIA, учитывающая все последовательности атакующих действий для устройства, вычисляется как максимальная оценка CIA по всем последовательностям атакующих действий для устройства.

Таблица 1

Группы уязвимостей

	Вектор атаки	**Требуемые привилегии**	**Полученные привилегии**
V0	Сетевая ИЛИ локальная	Нет	Нет
V1	Сетевая ИЛИ локальная	Нет	Нет
V2	Любая	Эквивалентно привилегиям V1	Любая
V3	Смежная сеть	Нет	Любая
V4	Любая	Эквивалентно привилегиям V3	Любая

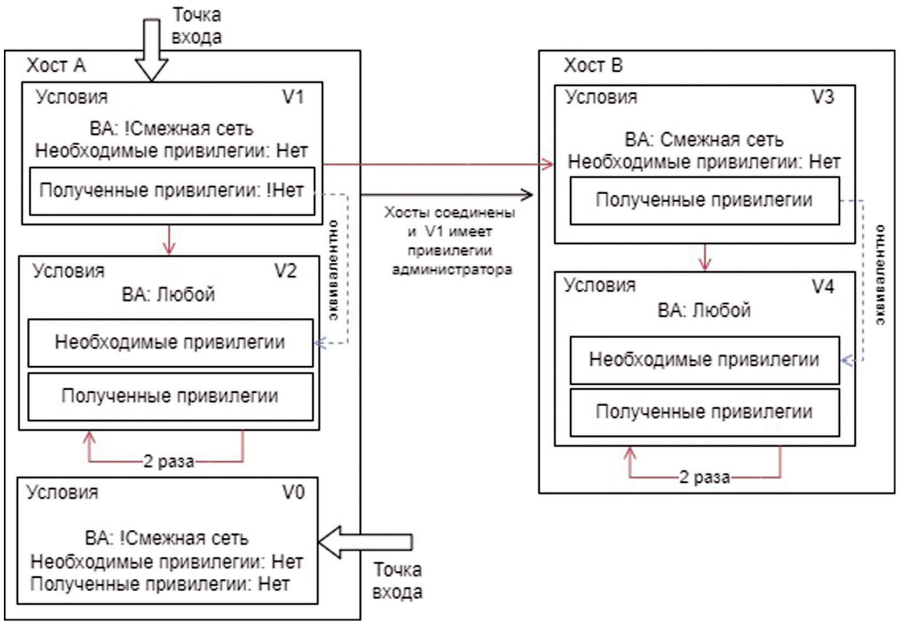

Рис. 4. Отношения между группами уязвимостей

Оценка CIA принимает значения в диапазоне [0; 10]. Численная оценка может быть преобразована в качественную оценку CIA в диапазоне {*низкий, средний, высокий*}.

3.6. Компонент вычисления динамических оценок CIA и рисков на основе статистических методов и методов машинного обучения

Компонент предназначен для обновления значений показателей рисков нарушения защищенности информации и конфиденциальности персональных данных на основе обнаруженных аномалий. Для обнаружения аномалий используются статистические методы и методы машинного обучения. В зависимости от количества обнаруженных аномалий рассчитываются весовые коэффициенты для оценки рисков, связанных с обработкой персональных данных, и оценки возможности успешной реализации атаки (рис.2). Оценка возможности успешной реализации атаки является динамической частью оценки CIA и поэтому также должна быть пересчитана.

На первом шаге для отдельного устройства или пользователя компонент строит векторы признаков путем агрегации данных о событиях на некотором временном интервале [14]. Затем компонент вычисляет диапазон значений признаков, соответствующих нормальной работе, для всех устройств или пользователей, и далее, используя эти диапазоны, проверяет всю активность пользователя или устройства в поиске значительных отклонений - аномалий.

Как было сказано выше, значения выбранных признаков вычисляются путем агрегирования на некотором временном интервале сообщений с учетом значений определенных атрибутов исходных логов (например, типа ошибок). Для оценки активности устройства на основе вычисленных характеристик в качестве рабочего был выбран временной интервал в 60 секунд. Эксперименты показали, что такого интервала достаточно для отслеживания некоторых незначительных изменений в активности устройства, и в то же время при использовании такого временного интервала сохраняется возможность обобщать закономерности в шаблонах активности устройства.

Для создания шаблона нормального поведения устройства и вычисления диапазона значений признаков, соответствующих нормальной работе устройства, были использованы следующие методы: вычисление интерквартильного размаха, тест Граббса, ESD-тест и экспоненциальное сглаживание. Значения, соответствующие нормальной работе устройства, используются в качестве входных данных для обнаружения аномалий.

Для обнаружения временных интервалов с аномалиями компонент выполняет следующие шаги:

(1) сравнивает значения признаков для каждого устройства с диапазонами нормальных значений для

обнаружения аномальной активности для устройств и вычисляет интенсивность аномалий, определяемую как отношение расстояния между значением признака и ближайшей границей диапазона к значению признака;

(2) использует метод машинного обучения Local Outlier Factor для обнаружения аномальной активности устройств;

(3) объединяет результаты с учетом интенсивности аномалии.

Были исследованы и другие методы машинного обучения, применяемые для обнаружения аномалий, такие как метод опорных векторов одного класса [15], алгоритм Isolation Forest [16], эллипсоидальная аппроксимация данных [17], искусственные нейронные сети различной структуры (автокодеры, LSTM, рекуррентные сети).

На последнем шаге работы компонента вычисляется количество временных интервалов с аномалиями разных типов для каждого устройства. Эта информация далее используется для вычисления весов аномалий и пересчета оценки рисков, связанных с обработкой персональных данных, а также оценки возможности успешной реализации атаки. Веса рассчитываются как отношение временных интервалов аномалий ко всем интервалам активности.

3.7. Компонент вычисления интегральных оценок рисков нарушения защищенности информации и конфиденциальности персональных данных

Компонент расчета интегральных показателей рисков нарушения защищенности информации и конфиденциальности персональных данных выполняет расчет обобщенной (интегральной) оценки рисков, связанных с использованием устройства. Применяется несколько подходов к вычислению интегральных показателей: экспертный (или табличный), минимаксный и подход на основе средневзвешенной суммы всех показателей.

Наиболее распространенным является подход на основе построения (экспертных) оценочных таблиц. В основном он используется для номинальных параметров. Первая строка и столбец такой таблицы содержат возможные значения входных показателей, а внутренние ячейки таблицы – значения интегральной оценки. Например, он используется в упрощенной процедуре анализа и оценки рисков (Facilitated Risk Analysis and Assessment Process, FRAAP), описанной в [18]. Очевидным преимуществом такого подхода является прозрачность процедуры расчета, однако создание таблиц для более чем трех показателей является достаточно сложным процессом.

Минимаксный подход обычно используется в задачах по выбору мер обеспечения безопасности и предполагает минимизацию таких параметров, как вероятность атаки, последствия атаки, затраты на реагирование, при условии максимизации таких показателей, как выгода от внедрения мер обеспечения безопасности [19].

Подход, основанный на вычислении средневзвешенной суммы, широко применяется для оценки рисков на основе числовых параметров, например, для расчета оценки уязвимостей по CVSS[16]. Использование средневзвешенной суммы требует определения рангов или весов для показателей. В некоторых случаях определение весов является вполне естественным процессом. Например, при вычислении интегральной статической оценки рисков на основе анализа политики безопасности персональных данных возможно учитывать два показателя – оценку удобочитаемости и оценку рисков на основе онтологии политик конфиденциальности. Оценка удобочитаемости характеризует простоту текста политики безопасности персональных данных, в то время как информация об использовании персональных данных включена в оценку, основанную на онтологии. Очевидно, что ее вес должен быть выше, чем вес оценки читаемости политики. Этот факт может быть отражен с помощью весовых коэффициентов. Авторы предлагают использовать следующие весовые коэффициенты для данных оценок: $w_o = 0.9$ для оценки на основе онтологии и $w_r = 0.1$ для оценки удобочитаемости политик конфиденциальности.

Алгоритм вычисления обобщенной оценки рисков обработки персональных данных *pps* на основе анализа политик будет иметь следующий вид. Пусть *os* – это оценка рисков, вычисленная на основе анализа онтологии. Значение оценки *os* находится в диапазоне [0, 10], где 0 соответствует нижней оценке рисков, 10 – верхней оценке рисков. Обозначим, через *rs* оценку удобочитаемости политики конфиденциальности, значения которой находятся в диапазоне [0, +∞).

Тогда алгоритм расчета интегральной оценки конфиденциальности персональных данных на основе политики включает следующие шаги:

1) **if** (rs > *rescaling_threshold*) **then** *rs* = *rescaling_threshold*.

2) *Rescaled_rs* = *rs* *10/*rescaling_threshold*.

3) $pps = w_o * os + w_r * rs$.

На шагах 1-2 алгоритма осуществляется приведение оценки читаемости в диапазон [0; 10]. Для это-

16 https://www.first.org/cvss/specification-document

го вводится пороговое значение *rescaling_threshold,* которое искусственно ограничивает возможное максимальное значение оценки удобочитаемости. Оно установлено равным 16, поскольку средний диапазон интервалов оценки удобочитаемости, определяющих разный уровень сложности текста, равен 4, а нижняя граница интервала, соответствующего самому высокому уровню сложности текста, равна 12.

Вычисляемая интегральная оценка рисков *pps* находится в диапазоне [0, 10].

Следующий небольшой пример иллюстрирует процедуру расчета интегральной оценки конфиденциальности на основе анализа политики конфиденциальности. Пусть оценка на основе анализа онтологии *os* равна 5.6, и оценка удобочитаемости *rs* равна 12, тогда интегральная оценка на основе анализа политики конфиденциальности равна:

1) $Rescaled\_rs = rs * 10/rescaling\_threshold = 12 * 10/16 = 7.5$.

2) $pps = w_o * os + w_r * rs = 0.9 * 5.6 + 0.1 * 7.5 = 5.8$.

В некоторых случаях определение весовых коэффициентов для показателей защищенности является нетривиальной задачей, и для решения этой проблемы авторы предлагают в качестве алгоритма расчета интегральной оценки риска защищенности и конфиденциальности подход, предложенный в [20]. Он предполагает, что все показатели являются одинаково значимыми. Показатель с наивысшей оценкой риска служит основой для вычисления оценки, вклад других показателей определяется логарифмической функцией от их значений.

Пусть *RiskScores* – множество оценок риска, для которых необходимо вычислить интегральную оценку. Тогда алгоритм вычисления интегральной оценки *IS* может быть описан следующим образом:

1. *max_score* = max(*RiskScores*).	Из множества имеющихся показателей риска выбрать показатель с максимальным значением, установить max_score равным его значению
2. *RiskScores* = *RiskScores* \ {*max_score*}.	Исключить его из множества показателей
3. $IS = max\_score + log\sum_i RiskScores$.	Вычислить интегральный показатель *IS* как сумму максимального значения оценки риска и логарифма суммы других оценок риска
4. **if** (*IS* > 10): *IS* = 10.	Если вычисленная оценка выше максимально возможного значения (10), установить ее равной 10

Использование логарифмической функции определяет нелинейный рост интегральной оценки и позволяет избежать быстрого увеличения ее значений. В [20] приведен детальный анализ чувствительности алгоритма к изменениям во входных параметрах.

Следует отметить, что все входные показатели должны иметь значения в диапазоне [0, 10], где 0 соответствует минимальному значению риска, а 10 – максимальному значению.

Результатом работы алгоритма является интегральная оценка в диапазоне [0, 10], где 0 соответствует минимальной оценке риска, а 10 – верхней максимальной оценке риска.

Ниже приведен пример расчета интегральной оценки защищенности и конфиденциальности *IS*. Пусть оценка рисков CIA равна 4.6, оценка рисков на основе прав доступа приложения APK равна 3.0, а оценка рисков на основе анализа политики конфиденциальности равна 4.5. В этом случае максимальной оценкой является оценка CIA, именно она берется за основу интегральной оценки:

IS = 4.6 + log(1 + (3.0 + 4.5), 10*2) = 5.2.

4. Экспериментальная оценка предложенного подхода

Обобщенная архитектура системы измерения защищенности представлена на рис. 5. Данная система была разработана на языке Python и использует PostgreSQL базу данных.

В базе данных представлена следующая информация: параметры анализируемых устройств, значения показателей, нормальные профили устройств, а также промежуточные данные об аномальной активности устройств. Кроме того, в базе хранятся данные об известных уязвимостях устройств Интернета вещей. При этом информация об уязвимостях обновляется автоматически с заданным интервалом. Также процедуру обновления можно запустить вручную.

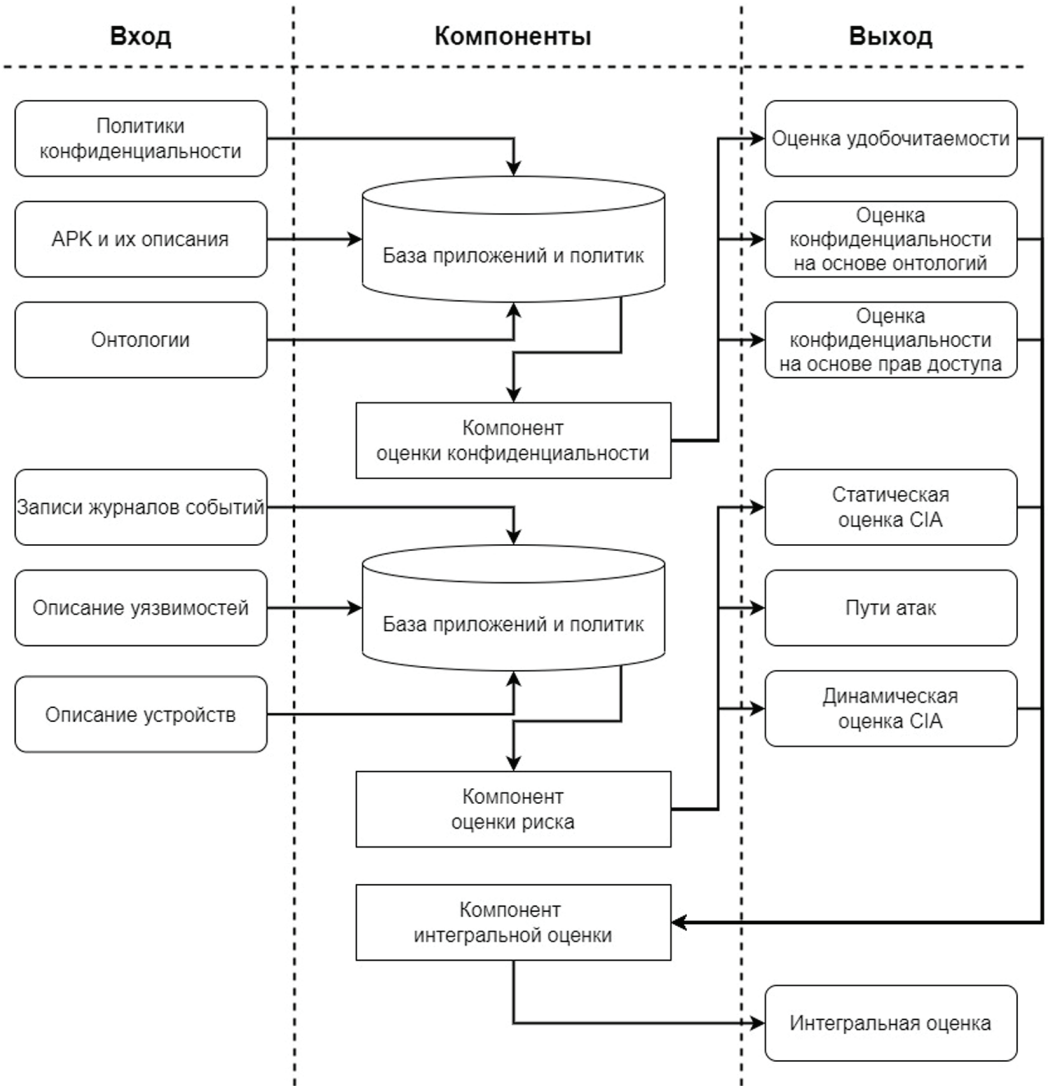

Рис. 5. Обобщенная архитектура системы измерения защищенности

Процесс работы системы измерения защищенности может быть представлен следующим образом. На начальном этапе система обновляет информацию об известных уязвимостях и принимает в качестве входных данных описание онтологии политик конфиденциальности. Затем пользователь заполняет спецификацию устройства, которая может включать тип устройства, модель, производителя и другую информацию, позволяющую идентифицировать устройство. При этом данная информация также может быть получена автоматически на основе данных системных журналов и журналов устройств. В дальнейшем, на основе собранной информации система автоматически ищет и загружает файлы приложений (APK, Android application package), а также их описание, доступное на торговых площадках приложений (например, Google Play). Кроме того, выгружается политика конфиденциальности устройств и/или производителя устройств. Альтернативно, данная информация может быть предоставлена пользователем системы вручную.

После обработки полученных данных система выводит статическую интегральную оценку риска нарушения защищенности и конфиденциальности. Для этого выполняются следующие шаги:

1. Вычисление оценки удобочитаемости для политики конфиденциальности устройства и/или производителя устройства.
2. Создание онтологии на основе заданного шаблона и оценка конфиденциальности ПДн на основе полученной онтологии.
3. Вычисление интегральной оценки конфиденциальности ПДн на основе используемой политики

(с учетом оценки удобочитаемости политики и оценки на основе онтологии).
4. Вычисление статической оценки рисков нарушения целостности, конфиденциальности и доступности на основе данных об уязвимостях, связанных с загруженными APK.
5. Вычисление оценки APK на основе их описаний и запрашиваемых прав доступа.
6. Вычисление интегральной оценки рисков, связанных с обработкой персональных данных, на основе APK.
7. Вычисление интегральной статической оценки рисков нарушения защищенности и конфиденциальности, а также ее представление пользователю.

Для получения динамических оценок, системе измерения защищенности необходим доступ к журнальным записям устройств, а также системы, в которой они работают. При этом система измерения защищенности извлекает из журнальных записей следующую информацию: данные о соединениях между устройствами; данные об устройствах, а именно характеристики устройств, их состояние, время входа и выхода из системы; данные об ошибках.

После этого система измерения защищенности вычисляет интегральную динамическую оценку защищенности и конфиденциальности. Для этого выполняются следующие шаги:

1. Обработка журнальных записей для генерации последовательностей атакующих действий на основе связей между устройствами и их уязвимостей.
2. Пересчет оценки рисков нарушения целостности, конфиденциальности и доступности (CIA) на основе сгенерированных последовательностей атакующих действий.
3. Обработка журнальных записей для выявления особенностей поведения устройства.
4. Создание профилей нормального поведения устройств на основе значений их характеристик.
5. Обнаружение аномальной активности на основе поступающих журнальных записей. Обработка новых журнальных записей осуществляется раз в сутки.
6. Вычисление динамической оценки рисков нарушения целостности, конфиденциальности и доступности (CIA) на основе обнаруженной аномальной активности.
7. Вычисление динамической оценки риска нарушения конфиденциальности ПДн на основе обнаруженной аномальной активности.
8. Вычисление интегральной динамической оценки риска нарушения защищенности и конфиденциальности, а также ее представление пользователю.

Все полученные результаты хранятся в разработанной базе данных и доступны пользователю.

Рассмотрим следующий пример функционирования системы измерения защищенности. Исследуемое устройство представляет собой умный замок производства компании August[17], занимающейся производством устройств для систем умного дома. Разработанный данной компанией умный замок позволяет реализовать следующие функции: удаленное открытие и закрытие дверей, журналирование входа и выхода пользователей в помещение, поддержка авторизации и голосового управления.

Для получения статической интегральной оценки защищенности и конфиденциальности в систему были загружены следующие данные:

— приложение August Home 11.5.1 для Android, которое управляет активностью устройств Умного дома, включая умный замок;
— описание August Home 11.5.1 для Android на торговой площадке APKPure;
— политика конфиденциальности с сайта производителя.

Системой была получена оценка удобочитаемости равная 13.0, что соответствует тексту высокой сложности. Также анализ политики конфиденциальности с использованием ее онтологического представления выявил подозрительные сценарии, такие как сбор персональных данных посетителей умного дома. На основе онтологии система получила оценку риска 8.37, что является довольно высоким показателем. При этом данный показатель может быть объяснен разнообразием типов собранной информации, которая собирается и передается: данные учетной записи пользователя, данные приложений и устройств, данные о платежной информации.

Таким образом, интегральная оценка риска на основе политики конфиденциальности составляет 8.34.

Анализ прав доступа и описания APK файла показал, что приложение August Home запрашивает следующие права доступа у пользователя:

— READ_CONTACTS (контакты),
— ACCESS_FINE_LOCATION (точное местоположение),
— ACCESS_COARSE_LOCATION (примерное местоположение),

17 https://august.com/pages/privacy-policy

— READ_EXTERNAL_STORAGE (внешняя память),
— READ_PHONE_STATE (состояние телефона, включая состояние сети, звонков, а также используемого аккаунта),
— CAMERA (камера) и
— RECORD_AUDIO (запись аудио).

При этом система на основе анализа политики конфиденциальности предположила запрос следующих прав доступа:

— ACCESS_FINE_LOCATION (точное местоположение),
— READ_EXTERNAL_STORAGE (внешняя память),
— READ_PHONE_STATE (состояние телефона, включая состояние сети, звонков, а также используемого аккаунта) и
— RECORD_AUDIO (запись аудио).

Это означает, что права доступа READ_CONTACTS (контакты), ACCESS_COARSE_LOCATION (примерное местоположение), а также CAMERA (камера) не были предсказаны системой. В результате были получены следующие результаты оценки риска на основе описаний и прав доступа:

— оценка на основе прав доступа $DS = 0.6$;
— оценка на основе описания $PS = 0.4$.

Таким образом, оценка риска на основе файла APK равна 7.1.

Дополнительно система измерения защищенности обнаружила следующую запись CPE в открытой базе уязвимостей NVD для анализируемого файла APK: cpe:2.3:a:august:august_home:-:*:*:*:*:android:*:*. Также была обнаружена соответствующая данной записи CPE уязвимость CVE: CVE-2019-17098. В соответствии с показателями CVSS данной уязвимости, а также оценкой критичности устройства, была получена статическая оценка рисков конфиденциальности, целостности и доступности, которая составляет 6.5.

При объединении статической оценки CIA с оценками на основе APK и политики конфиденциальности, система получает интегральную статическую оценку риска нарушения защищенности и конфиденциальности. Для рассмотренного примера данная оценка равна 9.2 при расчете с использованием нелинейного алгоритма и 7.4 при расчете с использованием функции взвешенной суммы.

Второй пример связан с вычислением динамических оценок. Вначале компонент обработки и интеграции записей из журнала событий системы измерения защищенности обрабатывает журналы событий. Входные журналы событий представляют собой файлы csv, разделенные по дням активности. Компонент объединяет и нормализует разнородные данные из журналов событий разных типов. Итоговый журнал содержит атрибуты времени регистрации события в журнале, связи между устройствами, информацию об устройстве (такую как идентификатор устройства, его модель, IP-адрес и др.), информацию о статусе устройства, время его входа/выхода из системы, и информацию об ошибках. После определения набора атрибутов для итогового журнала, сообщения, отсортированные по времени, помещаются в итоговый журнал. Вводится дополнительный атрибут – тип журнала, из которого было получено сообщение.

Итоговый журнал используется для генерации последовательностей атакующих действий на основе связей между устройствами и известных уязвимостей устройств. В описываемом примере на основе журнала событий система обнаружила связь между умным замком и устройством “August connect” (мост). В результате была сформирована последовательность из атакующего действия на умный замок, использующего уязвимость его APK (CVE-2019-17098) и атакующего действия на “August connect”, использующего его уязвимость (CVE-2018-20100). Компонент динамической оценки CIA с учетом последовательностей атакующих действий отнес CVE-2019-17098 к классу уязвимостей V3, а CVE-2018-20100 – к классу V1, в соответствии с рис.4. Поскольку CVE-2018-20100 позволяет получить права доступа администратора, можно сформировать следующую последовательность использования уязвимостей: CVE-2018-20100 -> CVE-2019-17098. В результате возможность использования CVE-2019-17098 с учетом последовательности увеличивается с 2.84 до 2.99. Оценка CIA, в свою очередь, увеличивается с 6.5 до 6.59, что не влияет на интегральную статическую оценку риска нарушения защищенности и конфиденциальности. Высокая интегральная статическая оценка риска нарушения защищенности и конфиденциальности растет достаточно медленно с ростом значения оценки CIA (оценка CIA должна увеличиться на 1.9, чтобы повлиять на интегральную оценку).

На следующем шаге система измерения защищенности обрабатывает журналы с целью вычисления признаков, описывающих поведение устройств, и формирования профилей нормального поведения устройств. Затем новые записи журналов событий обрабатываются раз в день с целью обнаружения аномалий. В случае их обнаружения динамическая оценка CIA и динамическая оценка конфиденциальности ПДн пересчитываются, что, в свою очередь, изменяет

интегральную динамическую оценку риска нарушения защищенности информации и конфиденциальности ПДн. Для валидации результатов вычисления динамической оценки в данной статье использовался сгенерированный журнал событий с аномалиями. Обнаруженные аномалии влияют на значение оценки CIA и оценки конфиденциальности ПДн. Так, в случае наличия в журнале 10% записей с аномалиями, оценка CIA увеличивается с 6.59 до 6.7. Но, как указано выше, оценка CIA должна увеличиться на 1.9, чтобы повлиять на интегральную оценку, и она остается равной 9.2.

5. Заключение

В статье описана разработанная система измерения защищенности для устройств Интернета вещей. Представленная система основана на оригинальной иерархии показателей защищенности и конфиденциальности, алгоритмах их расчета и алгоритмах расчета интегральных показателей. Разработанная система автоматизирует этапы сбора, обработки и анализа данных для расчета выбранных показателей и их пересчета при поступлении новых данных. Отличие предлагаемой системы от других аналогичных систем состоит в учете различных аспектов защищенности и конфиденциальности для сравнения уровня защищенности разных устройств Интернета вещей и объединении этих аспектов в единую интегральную оценку защищенности и конфиденциальности.

Лежащий в основе разработанной системы подход предполагает расчет базовой статической оценки устройства по его внутренним характеристикам и ее дальнейший пересчет в динамике с учетом новых данных, таких как данные о подключениях и поведении устройств. Это реализуется с помощью следующих шагов: вычисление оценки удобочитаемости политики конфиденциальности устройства/производителя; построение онтологии P2Onto на основе заданного шаблона и расчет оценки рисков нарушения конфиденциальности ПДн на основе онтологии; расчет интегральной оценки на основе политики конфиденциальности; расчет статической оценки CIA на основе уязвимостей, связанных с загруженными APK; расчет оценки на основе описания и оценки на основе прав доступа для APK; расчет интегральной оценки риска нарушения конфиденциальности на основе APK; расчет и вывод интегральной статической оценки риска нарушения защищенности и конфиденциальности; получение и обработка записей из журнала событий для построения профиля нормального поведения устройства; построение профиля нормального поведения устройства; получение и обработка записей из журнала событий каждый заданный интервал времени; перерасчет оценки CIA с учетом связей между устройствами и построение возможных последовательностей атакующих действий; проверка наличия аномалий в поведении устройства; пересчет оценок в случае наличия аномалии. Процесс продемонстрирован на примере устройства Интернета вещей.

В работе были предложены оригинальные алгоритмы для вычисления отдельных метрик и подход к их интеграции. Существуют различные подходы к расчету интегральных показателей [21], в том числе экспертный (или табличный) подход [17], минимаксный подход [19] и подход, основанный на взвешенной сумме[18]. Авторы выбрали подход взвешенной суммы, но модифицировали его, поскольку все входные показатели одинаково значимы. Алгоритм на основе взвешенной суммы уменьшил бы значение интегральной оценки риска нарушения защищенности и конфиденциальности, так как он усредняет значения. В случае оценок защищенности и конфиденциальности занижение оценок неприемлемо. В предлагаемом модифицированном алгоритме за основу берется показатель с наивысшим баллом, а к его значению добавляются значения других показателей. При этом вначале вычисляется логарифм, зависящий от их значений и максимальных значений для нелинейного масштабирования значения. Нелинейность вводится чтобы избежать быстрого роста значения интегрального показателя. Эксперименты показали, что предлагаемый алгоритм не уменьшает значение наибольшего показателя, так как оно выбирается в качестве базы интегральной оценки, и эта база увеличивается пропорционально значениям остальных показателей.

Ряд недостатков предлагаемой системы планируется устранить в дальнейшей работе. Могут быть добавлены новые показатели, такие как достоверность и прозрачность. В настоящее время значения критичности анализируемых устройств, используемые для расчета статической оценки CIA, устанавливаются в зависимости от типа устройства. В будущем расчет критичности может быть автоматизирован с учетом роли устройства в системе. Необходимо усовершенствовать автоматизированный поиск CPE и CVE, поскольку из-за отсутствия унификации названий программного и аппаратного обеспечения и ошибок в базе CPE иногда они пропускаются. Процесс обнару-

18 https://www.first.org/cvss/specification-document

жения аномалий можно улучшить, введя новые более сложные признаки. Кроме того, необходимо добавить обнаружение аномалий по профилю типа устройства. Также улучшить систему измерения защищенности может интеграция с системами обнаружения вторжений. Кроме того, в дальнейшей работе планируется добавить возможность формирования рекомендаций по повышению защищенности и конфиденциальности.

Кроме того, сложность состоит в валидации предложенной системы измерения защищенности. На данный момент эксперименты были проведены для ряда показателей, а применимость предложенной системы была показана на примере. Однако только результаты использования устройств и статистика реальных успешных инцидентов могут продемонстрировать корректность рассчитанных показателей защищенности и конфиденциальности. В дальнейшем планируется разработать тестовый стенд для оценки и компрометации устройств интернета вещей и провести тестирование системы на данных, полученных на стенде.

Рецензент: *Паращук Игорь Борисович, доктор технических наук, профессор, профессор Военной академии связи, Санкт-Петербург, Россия.*

E-mail: shchuk@rambler.ru

Литература

1. Doynikova E., Chechulin A., Kotenko I. Analytical attack modeling and security assessment based on the common vulnerability scoring system // Proceedings of the XXth Conference of Open Innovations Association FRUCT. — 2017. — P. 53–61. 10.23919/FRUCT.2017.8071292.
2. Wei R., Cai L., Yu A., Meng D. AGE: Authentication Graph Embedding for Detecting Anomalous Login Activities. — 2020. — DOI: 10.1007/978-3-030-41579-2_20.
3. Ardagna C.A., De Capitani di Vimercati S., Samarati P. Enhancing User Privacy Through Data Handling Policies // eds: Damiani E., Liu P., Proc. of the Data and Applications Security. – LNCS. — vol. 4127. — Springer, Berlin, Heidelberg, 2006.
4. Pardo R., Le Métayer D. Analysis of Privacy Policies to Enhance Informed Consent // Proc. of the Data and Applications Security and Privacy XXXIII (DBSec), eds.: Foley S., LNCS. - vol. 11559. - Springer, Cham, 2019.
5. Tesfay W.B., Hofmann P., Nakamura T., Kiyomoto S., Serna J. PrivacyGuide: Towards an Implementation of the EU GDPR on Internet Privacy Policy Evaluation // Proceedings of the Fourth ACM International Workshop on Security and Privacy Analytics (IWSPA '18). Association for Computing Machinery, New York, NY, USA, 2018. - P. 15–21. - DOI: https://doi.org/10.1145/3180445.3180447.
6. Kincaid J.P., Fishburne R.P., Rogers R.L., Chissom B.S. Derivation of new readability formulas (automated readability index, fog count, and flesch reading ease formula) for Navy enlisted personnel. Research Branch Report 8–75. Chief of Naval Technical Training: Naval Air Station Memphis, 1975.
7. Warsun N., Selo S., Widyawan W. Survey on Trust Calculation Methods in Internet of Things // Procedia Computer Science, 161. - 2019. - P. 1300–1307. - doi: 10.1016/j.procs.2019.11.245.
8. De S. J., Metayer D. L. Privacy Risk Analysis to Enable Informed Privacy Settings // Proc. of the 2018 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), London, 2018. P. 95–102.
9. Bar-Sinai M., Sweeney L., Crosas M. DataTags, Data Handling Policy Spaces and the Tags Language // Proc. of the 2016 IEEE Security and Privacy Workshops (SPW), San Jose, CA, 2016. - P. 1–8.
10. Metayer D.L. A Formal Privacy Management Framework // eds.: Degano P., Guttman J., Martinelli F., Proc. of the Formal Aspects in Security and Trust (FAST), 2008, LNCS. - vol. 5491. - Springer, Berlin, Heidelberg, 2009.
11. Pandit H.J., Fatema K., O'Sullivan D., Lewis D.: GDPRtEXT - GDPR as a Linked Data Resource // eds.: Gangemi A. et al., Proc. of The Semantic Web (ESWC), 2018, LNCS. - vol. 10843. - Springer, Cham, 2018.
12. Ushakov R., Doynikova E., Novikova E., Kotenko I. CPE and CVE based Technique for Software Security Risk Assessment // The 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS 2021). - Cracow, Poland, 2021. - P. 353-356. - DOI:10.1109/IDAACS53288.2021.9660968.
13. Novikova E., Doynikova E., Kotenko I. P2Onto: Making Privacy Policies Transparent // The 3rd International Workshop on Attacks and Defenses for Internet-of-Things (ADIoT 2020), In Conjunction with ESORICS 2020. 4-6 November 2020, Paris, France. / Computer Security, Lecture Notes in Computer Science (LNCS). – Springer, 2020. - vol. 12501. - P. 235-252. - DOI: https://doi.org/10.1007/978-3-030-64330-0_15.
14. Котенко И.В., Федорченко А.В., Саенко И.Б., Кушнеревич А.Г. Технологии больших данных для корреляции событий безопасности на основе учета типов связей // Вопросы кибербезопасности. 2017. № 5 (23). С.2-16. DOI: 10.21681/2311-3456-2017-5-2-16.
15. Schölkopf B., Platt J. C., Shawe-Taylor J., Smola A. J., Williamson R. C. Estimating the Support of a High-Dimensional Distribution // Neural Computation, 13, 7, 2001. - P. 1443–1471. - doi:10.1162/089976601750264965.
16. Rousseeuw P.J., Van Driessen K. A fast algorithm for the minimum covariance determinant estimator // Technometrics, 41, 3, 1999. - P. 212.
17. Liu F. T., Ting K. M., Zhou Z.-H. Isolation-Based Anomaly Detection // ACM Transactions on Knowledge Discovery from Data, 6, 1, 2012. - P. 1–39. - doi:10.1145/2133360.2133363.
18. Peltier T.R. Information security risk analysis, 3d edition, CRC Press, 2010, 456 p.
19. Khouzani M.H.R., Liu Z., Malacaria P. Scalable min-max multi-objective optimization over probabilistic attack graphs // European Journal of Operational Research, vol. 278. - issue 3. - 2019. - P. 894–903.
20. Novikova E., Doynikova E., Gaifulina D., Kotenko I. Construction and Analysis of Integral User-Oriented Trustworthiness Metrics // Electronics. – 2022. – 11(2): 234. - https://doi.org/10.3390/electronics11020234.
21. Оценивание защищенности и выбор контрмер для управления кибербезопасностью. Монография / Е.В. Дойникова и И.В. Котенко. СПб.: Изд-во «Наука», 2021. – 197 с. ISBN 978-5-907366-23-7.

THE SECURITY AND PRIVACY MEASURING SYSTEM FOR THE INTERNET OF THINGS DEVICES

Elena Fedorchenko[19], Evgenia Novikova[20], Igor Kotenko[21], Diana Gaifulina[22], Olga Tushkanova[23], Dmitry Levshun[24], Alexey Meleshko[25], Ivan Murenin[26], Maxim Kolomeec[27]

The purpose of the article: *elimination of the gap in existing need in the set of clear and objective security and privacy metrics for the IoT devices users and manufacturers and an absence of such a set incorporating the interconnected security and privacy metrics, the algorithms for their calculation and generation of the integral clear and objective score by the development of the security and privacy measuring system for the IoT devices.*

Research method: *theoretical and system analysis for determination and classification of the security and privacy metrics, semantic analysis for generating of the semantic model of personal data processing scenarios, analytical modeling methods for generating of the attack traces, log analysis methods, statistical methods and machine learning methods for searching of the anomalies in device behavior, development of the database and software implementing the proposed security and privacy measuring system.*

The result obtained: *the security and privacy measuring system for the IoT devices users and manufacturers is proposed. The proposed system allows automated calculation of the security and privacy metrics based on the available data on the device and generation of the integral security and privacy score. The hierarchy of security and privacy metrics is developed in the scope of the proposed system. The proposed metrics are calculated using static and dynamic data on the device and its behavior. Original algorithms for calculation of the outlined metrics are developed, including the algorithms for calculation of the integral security and privacy score. The architecture of the security measuring system is developed. It integrates the components implementing the developed algorithms for metrics calculation. The system operation is demonstrated on the case study.*

The area of use of the proposed approach - the developed security and privacy measuring system can be used by the IoT devices manufacturers to analyse their security and privacy, and to provide the users with simple and clear security and privacy metrics.

Novelty: *the hierarchy of static and dynamic security and privacy metrics for the Internet of Things is developed; the approach to security and privacy assessment for the Internet of Things on the basis of the developed metrics and available data is proposed; novel algorithms for metrics calculation are developed; novel algorithms for integral metrics calculation considering available data are developed.*

Contribution: *Fedorchenko E. – development of the approach, metrics hierarchy, and system architecture, problem statement for the components and their development, Novikova E. – the component for calculation of privacy risks, the component for calculation of integral risk scores, Kotenko I. – project management, problem*

19 Elena V. Fedorchenko (Doynikova), Ph.D, Senior researcher at St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), St. Petersburg, Russia. E-mail: doynikova@comsec.spb.ru

20 Evgenia S. Novikova, Ph.D, Senior researcher at St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), St. Petersburg, Russia. E-mail: novikova@comsec.spb.ru

21 Igor V. Kotenko, Dr.Sc., Professor, chief Scientist and Head of Laboratory of Computer Security Problems at St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), St. Petersburg, Russia. E-mail: ivkote@comsec.spb.ru

22 Gaifulina A. Diana, post-graduate student, junior researcher, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), St. Petersburg, Russia. E-mail: gaifulina@comsec.spb.ru

23 Olga N. Tushkanova, Ph.D, Senior researcher at St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), St. Petersburg, Russia. E-mail: tushkanova@comsec.spb.ru

24 Dmitry S. Levshun, Candidate of Technical Sciences, PhD in Computer Science, Senior Researcher, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), St. Petersburg, Russia. E-mail: levshun@comsec.spb.ru

25 Meleshko V. Alexey, post-graduate student, junior researcher, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), St. Petersburg, Russia. E-mail: meleshko.a@iias.spb.su

26 Ivan N. Murenin, PhD student, junior researcher, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), St. Petersburg, Russia. E-mail: imurenin@gmail.com

27 Kolomeets V. Maxim, PhD in Computer Science, senior researcher, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), St. Petersburg, Russia. E-mail: kolomeec@comsec.spb.ru

*statement, system architecture, Gaifulina D. – the component for event logs processing and integration, Tushkanova O., Murenin I. – the component for calculation of the dynamic risks score using statistical methods and machine learning, Levshun D. – metrics database, the component for calculation of the static risk score, Meleshko A. – the component for readability assessment, Kolomeets M. – the component for privacy risks assessment on the basis of *.apk files, the component for the dynamic risk score calculation considering attacks traces. All authors participated in the writing of the article.*

Keywords: *security assessment, internet of things, metrics, integral scores, static information, dynamic information, semantics, ontology, cyber attack, information system, data mining, anomalies, attack traces.*

References

1. Doynikova E., Chechulin A., Kotenko I. Analytical attack modeling and security assessment based on the common vulnerability scoring system // Proceedings of the XXth Conference of Open Innovations Association FRUCT, 2017. P. 53–61. 10.23919/FRUCT.2017.8071292.
2. Wei R., Cai L., Yu A., Meng D. AGE: Authentication Graph Embedding for Detecting Anomalous Login Activities, 2020. doi: 10.1007/978-3-030-41579-2_20.
3. Ardagna C.A., De Capitani di Vimercati S., Samarati P. Enhancing User Privacy Through Data Handling Policies // eds: Damiani E., Liu P., Proc. of the Data and Applications Security, 2006, LNCS, vol. 4127, Springer, Berlin, Heidelberg.
4. Pardo R., Le Métayer D. Analysis of Privacy Policies to Enhance Informed Consent // Proc. of the Data and Applications Security and Privacy XXXIII (DBSec), eds.: Foley S., LNCS, vol. 11559, Springer, Cham, 2019.
5. Tesfay W.B., Hofmann P., Nakamura T., Kiyomoto S., Serna J. PrivacyGuide: Towards an Implementation of the EU GDPR on Internet Privacy Policy Evaluation // Proceedings of the Fourth ACM International Workshop on Security and Privacy Analytics (IWSPA '18), Association for Computing Machinery, New York, NY, USA, 2018. P. 15–21. doi: https://doi.org/10.1145/3180445.3180447.
6. Kincaid J.P., Fishburne R.P., Rogers R.L., Chissom B.S. Derivation of new readability formulas (automated readability index, fog count, and flesch reading ease formula) for Navy enlisted personnel. Research Branch Report 8–75. Chief of Naval Technical Training: Naval Air Station Memphis, 1975.
7. Najib Warsun, Sulistyo Selo, Widyawan Widyawan. Survey on Trust Calculation Methods in Internet of Things // Procedia Computer Science, 161, 2019. P. 1300–1307. doi: 10.1016/j.procs.2019.11.245.
8. De S. J., Metayer D. L. Privacy Risk Analysis to Enable Informed Privacy Settings // Proc. of the 2018 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), London, 2018. P. 95–102.
9. Bar-Sinai M., Sweeney L., Crosas M. DataTags, Data Handling Policy Spaces and the Tags Language // Proc. of the 2016 IEEE Security and Privacy Workshops (SPW), San Jose, CA, 2016. P. 1–8.
10. Metayer D. L. A Formal Privacy Management Framework // eds.: Degano P., Guttman J., Martinelli F., Proc. of the Formal Aspects in Security and Trust (FAST), 2008, LNCS, vol. 5491, Springer, Berlin, Heidelberg, 2009.
11. Pandit H.J., Fatema K., O'Sullivan D., Lewis D.: GDPRtEXT - GDPR as a Linked Data Resource // eds.: Gangemi A. et al., Proc. of The Semantic Web (ESWC), 2018, LNCS, vol. 10843, Springer, Cham, 2018.
12. Roman Ushakov, Elena Doynikova, Evgenia Novikova, Igor Kotenko. CPE and CVE based Technique for Software Security Risk Assessment // The 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS 2021), 2021. 22-25 September, 2021, Cracow, Poland. P. 353-356. DOI:10.1109/IDAACS53288.2021.9660968.
13. Novikova E., Doynikova E., Kotenko I. P2Onto: Making Privacy Policies Transparent // The 3rd International Workshop on Attacks and Defenses for Internet-of-Things (ADIoT 2020), In Conjunction with ESORICS 2020. 4-6 November 2020, Paris, France. / Computer Security, Lecture Notes in Computer Science (LNCS), Springer. 2020. vol. 12501 LNCS. pp. 235-252. DOI: https://doi.org/10.1007/978-3-030-64330-0_15 (WoS, Scopus) eLIBRARY ID: 45049659.
14. Kotenko I., Fedorchenko A., Saenko I., Kushnerevich A. Big data technologies for of security event correlation based on type accounting // Cybersecurity issues. 2017. No. 5 (23). C.2-16. DOI: 10.21681/2311-3456-2017-5-2-16.
15. Schölkopf B., Platt J. C., Shawe-Taylor J., Smola A. J., Williamson R. C. Estimating the Support of a High-Dimensional Distribution // Neural Computation, 13, 7, 2001. P. 1443–1471. doi:10.1162/089976601750264965.
16. Rousseeuw P.J., Van Driessen K. A fast algorithm for the minimum covariance determinant estimator // Technometrics, 41, 3, 1999. P. 212.
17. Liu F. T., Ting K. M., Zhou Z.-H. Isolation-Based Anomaly Detection // ACM Transactions on Knowledge Discovery from Data, 6, 1, 2012. P. 1–39. doi:10.1145/2133360.2133363.
18. Peltier T.R. Information security risk analysis, 3d edition, CRC Press, 2010, 456 p.
19. Khouzani MHR., Liu Z., Malacaria P. Scalable min-max multi-objective optimization over probabilistic attack graphs // European Journal of Operational Research, vol. 278, issue 3, 2019. P. 894–903.
20. Novikova E., Doynikova E., Gaifulina D., Kotenko I. Construction and Analysis of Integral User-Oriented Trustworthiness Metrics // Electronics. – 2022. – 11(2): 234. - https://doi.org/10.3390/electronics11020234.
21. Security assessment and selection of countermeasures for cybersecurity management. Monograph / E.V. Doinikov and I.V. Kotenko. St. Petersburg: Nauka Publishing House, 2021. - 197 p. ISBN 978-5-907366-23-7.