УДК 004.056.5: 004.822

Методика оценки актуальных угроз и уязвимостей на основе технологий когнитивного моделирования и Text Mining

Васильев В. И., Вульфин А. М., Кириллова А. Д., Кучкарова Н. В.

Постановка задачи: *автоматизация низкоуровневого моделирования сценариев эксплуатации уязвимостей и реализации угроз на основе описания шаблонов компьютерных атак, содержащихся в базах данных, характеризующих различные аспекты безопасности программного и аппаратного обеспечения инфраструктуры сети промышленных объектов.* ***Целью работы*** *является разработка методики автоматизированной оценки актуальных угроз и уязвимостей согласно методике Федеральной службы по техническому и экспортному контролю (ФСТЭК) России.* ***Используемые методы****: семантическое моделирование на основе перекрестных ссылок и оценке степени сходства текстовых описаний объектов промышленной сети, заимствованных из банка данных угроз безопасности информации (БДУ) ФСТЭК России и баз CAPEC, MITRE и ATT&CK. Моделирование основано на формализации логической цепочки: «множество выявленных уязвимостей программного обеспечения; множество релевантных угроз; множество наиболее вероятных сценариев реализации угроз; возможные киберфизические последствия; количественная оценка рисков нарушения кибербезопасности» с учетом требований нормативных документов ФСТЭК России. Формализация выполнена в нотации семантических моделей и нечетких когнитивных моделей.* ***Новизна:*** *семантическая модель расширяет ссылочную модель отношений аспектов безопасности программного и аппаратного обеспечения. Отличие модели заключается в использовании весовых коэффициентов на основе оценки степени близости векторов вложений Doc2Vec текстовых описаний объектов. Использование модели при выполнении основных этапов анализа согласно методике ФСТЭК России позволяет снизить трудоемкость формирования перечня актуальных угроз и уязвимостей за счет применения технологий Text Mining для префильтрации несвязанных или недостижимых вершин. Построение когнитивной модели оценки рисков нарушения кибербезопасности для целевых объектов автоматизированной системы управления технологическими процессами (АСУ ТП) на основе семантической модели позволяет получить детализированную количественную оценку рисков нарушения кибербезопасности и сделать более обоснованный выбор средств защиты информации за счет моделирования различных сценариев реализации угроз.* ***Результат****: представлена методика оценки актуальных (т.е. потенциально наиболее опасных) уязвимостей и угроз программного обеспечения АСУ ТП. Рассмотрен пример применения предложенной методики для оценки угроз и уязвимостей прикладного программного обеспечения подсистемы АСУ ТП промышленного объекта нефтедобычи с последующей количественной оценкой рисков нарушения кибербезопасности.* ***Практическая значимость:*** *предлагаемый подход позволяет автоматизировать низкоуровневое моделирование сценариев реализации угроз на основе описания шаблонов компьютерных атак и снизить трудоемкость формирования перечня актуальных угроз за счет применения технологий Text Mining при выполнении основных этапов анализа согласно методике ФСТЭК России.*

Ключевые слова: *уязвимость; угроза; кибератака; системы классификации; семантический анализ текстов.*

Библиографическая ссылка на статью:
Васильев В. И., Вульфин А. М., Кириллова А. Д., Кучкарова Н. В. Методика оценки актуальных угроз и уязвимостей на основе технологий когнитивного моделирования и Text Mining // Системы управления, связи и безопасности. 2021. № 3. С. 110-134. DOI: 10.24412/2410-9916-2021-3-110-134

Reference for citation:
Vasilyev V. I., Vulfin A. M., Kirillova A. D., Kuchkarova N. V. Methodology for Assessing Current Threats and Vulnerabilities Based on Cognitive Modeling Technologies and Text Mining. *Systems of Control, Communication and Security*, 2021, no. 3, pp. 110-134 (in Russian). DOI: 10.24412/2410-9916-2021-3-110-134

Введение

Стремительное внедрение цифровых технологий во всех сферах экономики сегодня сопровождается таким же бурным ростом числа компьютерных атак на значимые информационные ресурсы государственных и коммерческих организаций, промышленных предприятий, юридических и частных лиц. Так, согласно данным, приведенным в аналитическом отчете компании Positive Technologies [1], в IV квартале 2020 г. был зафиксирован рост инцидентов на 3,1%, по сравнению с III кварталом и на 41,2% по сравнению с аналогичным периодом 2019 г. Целенаправленные атаки составили 80% от общего числа атак. В поле зрения злоумышленников при этом часто попадают сетевые ресурсы промышленных компаний, доступные из Интернета. В IV квартале треть всех инцидентов информационной безопасности (ИБ) в промышленности связаны с эксплуатацией уязвимостей и недостатков защиты информационных систем (ИС), в 84% атак применялось вредоносное программное обеспечение (ПО). По данным Лаборатории Касперского [2], в 2019 г. было выявлено 509 новых уязвимостей в различных компонентах автоматизированной системы управления технологическими процессами (АСУ ТП), более половины которых получили оценку более 7 баллов по шкале CVSS 3.0, что соответствует высокой и критической степени риска.

Очевидно, что в такой ситуации специалисты по ИБ, ответственные за безопасное функционирование промышленных автоматизированных систем, остро нуждаются в наличии эффективной информационно-аналитической поддержки, с помощью которой они могли бы не только своевременно реагировать на ту или иную реальную атаку, но и прогнозировать (предупреждать) появление такого рода атак. Сегодня подобная поддержка во многом становится реальностью. В обиход специалиста по ИБ на наших глазах входят такие новые понятия, как базы знаний угроз (Threat Intelligence) [3, 4] и уязвимостей (Vulnerability Intelligence) [5], под которыми понимаются полученные из различных источников и соответствующим образом систематизированные сведения, касающиеся методов, используемых нарушителями для нанесения ущерба, механизмов использования уязвимостей, возможных последствиях от реализации атак и способах противодействия им.

На сегодняшний день систематизированные текстовые описания, характеризующие различные аспекты безопасности программного и аппаратного обеспечения информационной инфраструктуры (дескрипторы безопасности), консолидированы в виде слабосвязанных групп иерархических гипертекстовых документов в отдельных базах данных [6-9]:

- CVE (Common Vulnerabilities and Exposures) – база данных (стандарт) в области унификации именования и регистрации обнаруженных уязвимостей ПО;
- CWE (Common Weakness Enumeration) – база данных недостатков (слабых мест) ПО, которые могут быть использованы нарушителями при проведении атак;
- NDV (NIST National Vulnerability Database) – представительная база данных уязвимостей ПО;

- CVSS (Common Vulnerability Scoring System) – система рейтинговой оценки опасности уязвимостей ПО;
- CAPEC (Common Attack Pattern Enumeration and Classification) – стандарт описания классов атак и их иерархических отношений, каталог известных кибератак;
- MITRE ATT&CK Matrix – формальное описание техник и тактик реализации кибератак;
- CPE (Common Platform Enumeration) – формальный язык описания всех возможных продуктов, операционных систем и аппаратных устройств при описании уязвимостей;
- Банк данных угроз безопасности информации (БДУ) ФСТЭК России – база данных уязвимостей и угроз.

В феврале 2021 г. утвержден методический документ ФСТЭК России «Методика оценки угроз безопасности информации» [10], который определяет порядок моделирования и оценки актуальности угроз и возможные сценарии их реализации в информационных системах на основе БДУ ФСТЭК и перечисленных выше баз данных компьютерных атак.

В то же время, возможность открытого доступа к подобной информации еще сама по себе не гарантирует ее эффективность, поскольку специалист по ИБ пока вынужден «вручную» справляться с огромным объемом данных. Так, база данных NDV по состоянию на начало 2021 г. содержит более 150 тыс. записей об уязвимостях, а БДУ ФСТЭК России – около 30 тыс. записей об уязвимостях и 222 записи об угрозах безопасности информации. Указанная информация хранится в виде текстовых описаний, анализ которых требует существенных временных затрат и определенных профессиональных навыков. Поэтому понятен тот интерес, который сегодня проявляется к использованию методов семантического (интеллектуального) анализа текстов [11-13], применение которых позволило бы решить в какой-то степени проблему автоматизации поиска и анализа необходимой специалисту конкретной информации в перечисленных выше источниках данных. Целью является разработка методики автоматизированной оценки актуальных угроз и уязвимостей ПО АСУ ТП на основе технологий когнитивного моделирования и Text Mining с возможностью количественной оценки рисков нарушения кибербезопасности, расширяющей методику ФСТЭК России.

Анализ семантических моделей формализации знаний о дескрипторах безопасности на основе технологий Text Mining

Работа [14] посвящена задаче ранжирования (приоритезации) уязвимостей ПО, для решения которой предложена следующая процедура: а) формируется корпус CVE-описаний уязвимостей из базы данных NVD; б) выполняется конвертация CVE-описаний в «мешок слов» (модель CBOW для представления текстов); в) осуществляется ранжирование уязвимостей по степени важности с применением алгоритма TextRank, основанного на оценке семантической близости предложений в тексте. В [15] решается задача поиска подходящих шаб-

лонов атак в базе данных CAPEC по запросу пользователя (специалиста по ИБ). При этом, в соответствии с технологией Text Mining, шаблоны атак рассматриваются как элементы многомерного векторного пространства, в котором с помощью меры семантической близости осуществляется кластеризация шаблонов атак. Результаты, в свою очередь, отображаются в виде соответствующих диаграмм.

В работе [16] предложено связать тактики и техники из MITRE ATT&CK Matrix, CWE, CVE и список CAPEC на основе существующих в гипертекстовых документах перекрестных ссылок. Сохраняются все дескрипторы безопасности объектов ИС и взаимосвязи источников, одновременно обеспечивается двунаправленная реляционная связь в результирующем графе. Ограничения подхода: используются только данные NVD, и не учитывается множество уязвимостей из других баз и уязвимостей, не имеющих идентификаторов CVE. Данные для построения взяты из разных источников и не всегда удается сохранить непротиворечивость и полноту результирующего графа.

В статье [17] предлагается объединить дескрипторы из отдельных баз данных (CVE, CWE, CAPEC) в согласованный граф знаний. Предложен метод построения графа знаний, который включает не только реляционную, но и текстовую информацию о дескрипторах безопасности в непрерывном векторном пространстве, развивая методы построения вложений для текстовых данных (Word2Vec).

Для взаимодействия с иерархией классов атак, представляющих текстовые, категориальные и ссылочные данные в документах CAPEC, в работе [18] применяется визуализация двудольного графа ссылок на шаблоны атак на дескрипторы CWE.

NVD предоставляет возможность экспорта описаний уязвимостей в машиночитаемых форматах. Часть записей CVE не содержат меток Common Platform Enumeration (CPE) -version, -product и -vendor, что затрудняет автоматическое обнаружение уязвимостей. В работе [19] предложена процедура автоматического сопоставления перечней CVE с CPE с помощью алгоритмов машинного обучения, реализующих методы распознавания именованных сущностей (NER).

В статье [20] на основе методов машинного обучения и обработки естественного языка выполняется прогнозирование последствий кибератак. Предлагаемые инструменты упрощают информирование о последствиях атаки и позволяют снизить когнитивную нагрузку на исследователей за счет автоматического прогнозирования последствий в случае обнаружения новых атак.

Таким образом, предпринимаются активные усилия по разработке методов и подходов к построению семантических моделей для формализации знаний о дескрипторах безопасности объектов ИС на основе технологий Text Mining.

В последующих разделах данной статьи авторы на основе семантического сходства текстовых описаний, заимствованных из БДУ ФСТЭК России и баз CAPEC и MITRE ATT&CK, рассматривают вопросы построения и формализации логической цепочки: «множество выявленных уязвимостей ПО → множество релевантных угроз → множество наиболее вероятных сценариев реализа-

ции угроз → возможные киберфизические последствия → количественная оценка рисков нарушения кибербезопасности» с учетом требований нормативных документов ФСТЭК. Формализация логических отношений выполнена в нотации семантических моделей и нечетких когнитивных моделей.

Методика оценки актуальных угроз и уязвимостей ПО АСУ ТП с использованием методов семантического анализа дескрипторов безопасности

«Методика оценки угроз безопасности информации» ФСТЭК России [10] определяет порядок моделирования и оценки актуальности угроз для всех типов ИС на основе перечня актуальных уязвимостей. Функциональная модель оценки угроз и сценариев их реализации на основе перечня уязвимостей представлена на рис. 1.

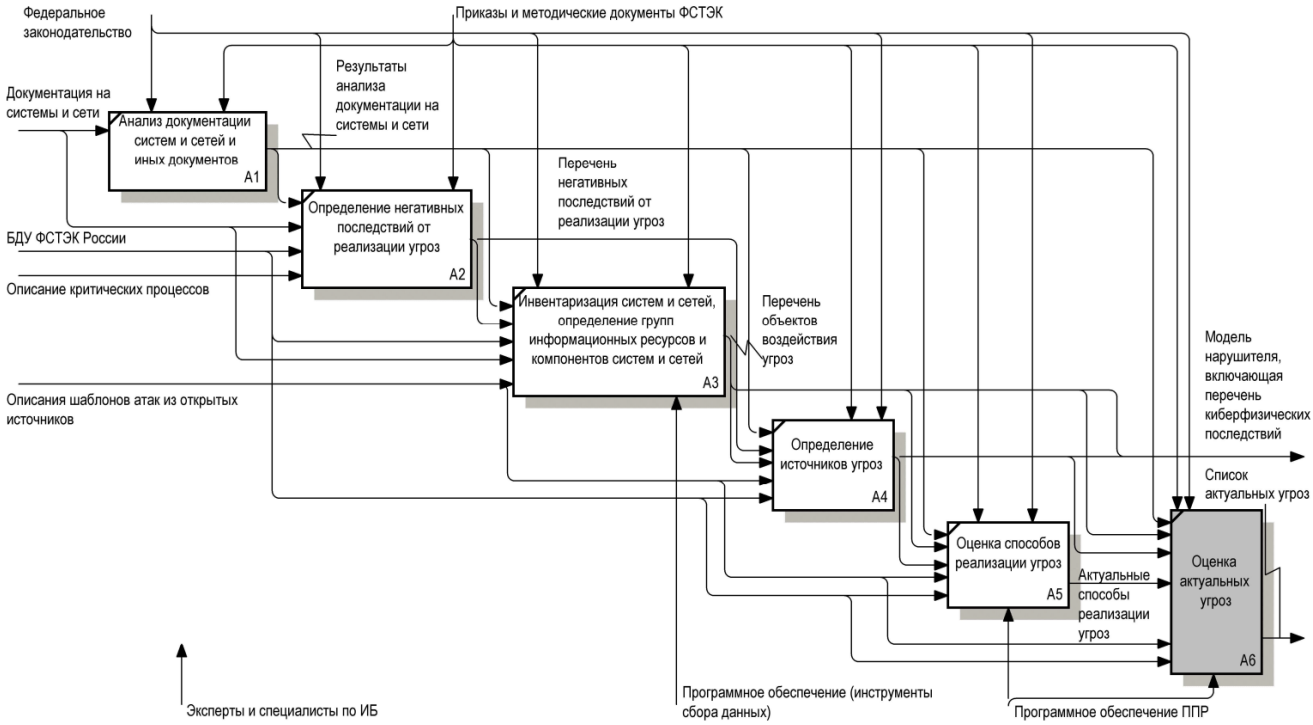


Рис. 1. Функциональная модель оценки угроз и уязвимостей

Методика объединяет стратегический (определение нарушителя, цели воздействия, основные негативные последствия) и тактический (применяемые тактики и техники эксплуатации уязвимостей, образующие возможные сценарии реализации угроз) уровни построения модели угроз. Основные затруднения при использовании текущей редакции методики заключаются в трудоемкости разработки сценариев [21] (декомпозиция блоков А5 и А6 на рис. 1) и их слабой связанности с сформированными на предыдущих шагах (А1-А4 на рис. 1) перечнями актуальных уязвимостей, типов доступа, типов нарушителей, видов ущерба, целей и т.п., а также в отсутствии инструментов поддержки принятия решение и моделирования сценариев реализации угроз.

Следовательно, актуальным является автоматизация низкоуровневого моделирования сценариев эксплуатации уязвимостей и реализации угроз на осно-

ве описания шаблонов компьютерных атак, содержащихся в базах данных и иных источниках, опубликованных в сети «Интернет» (CAPEC, ATT&CK, OWASP, STIX, WASC и др.). Для этого предлагается обобщить рассмотренные ранее подходы к построению семантических моделей дескрипторов безопасности объектов ИС (рис. 2), рассмотренные в [22], методы оценки актуальных угроз и уязвимостей на основе технологий Text Minining, и технологии когнитивного моделирования, предложенные в [23, 24].

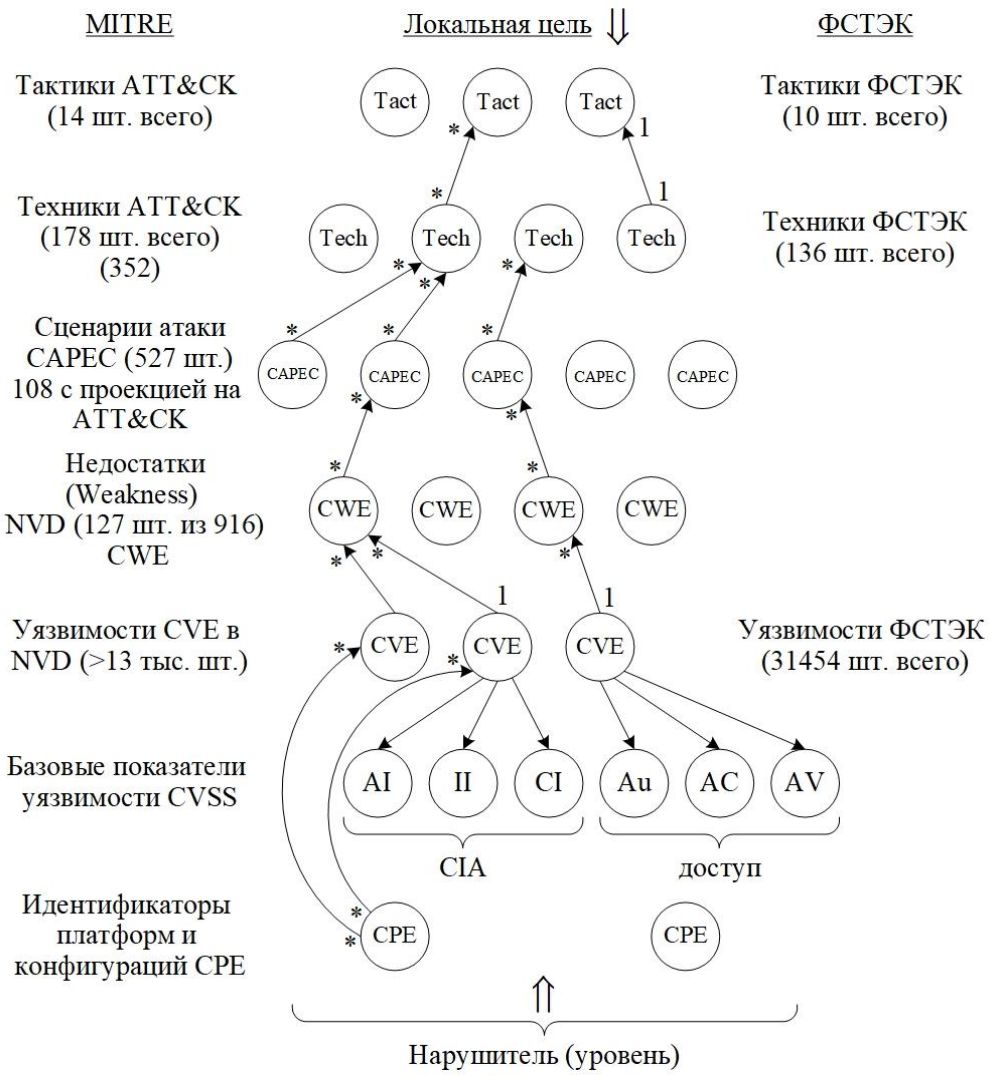


Рис. 2. Семантическая модель дескрипторов безопасности объектов ИС в задаче анализа актуальных угроз и уязвимостей

На рис. 2 представлены основные уровни детализации шаблона атаки, охватывающие описание выявленных и потенциальных уязвимостей, сценариев, тактик и техник ее реализации. Семантическая модель представляет собой граф:

$$G = \{V, R, D\},$$

где V – множество вершин графа – дескрипторов безопасности;

$$V = V_{CPE} \cup V_{CVE} \cup V_{CWE} \cup V_{CAPEC} \cup V_{Techs} \cup V_{Tackts},$$

V_{CPE} – множество вершин, соответствующих идентификаторам платформ и конфигураций для программно-аппаратного обеспечения ИС;

V_{CVE} – множество вершин, соответствующих идентификаторам выявленных уязвимостей для каждого компонента ИС;

V_{CWE} – множество вершин, соответствующих дескрипторам CWE, представляющим недостатки (слабые места) программного и аппаратного обеспечения ИС;

V_{CAPEC} – множество вершин, соответствующих меташаблонам CAPEC, описывающим известные типовые атаки;

V_{Techs} – множество вершин, соответствующих техникам реализации атаки, которые описывают инструменты, технологии, утилиты и т.д., используемые нарушителями;

V_{Tackts} – множество вершин, соответствующих тактикам, т.е. действиям на разных этапах реализации атаки;

R – множество взвешенных ориентированных ребер, устанавливающих отношения между дескрипторами:

$$R \subseteq V \times V,\ r\left(v_i, v_j\right),\ v_i, v_j \in V,$$

$D\left(r\right)$ – функция, определяющая степень семантической близости для концептов $v_i, v_j \in V$.

Семантическая модель G дескрипторов безопасности объектов ИС строится на основе перекрестных ссылок (r) в гипертекстовых документах (дескрипторах безопасности объектов ИС – V графа) с поддержкой основных типов отношений между сущностями в нотации реляционных моделей (отношения «один ко многим», «один к одному», «многие ко многим») и в нотации UML диаграмм вариантов использования для описания типов связей «является потомком», «является родителем», «может предшествовать» и т.п. Дополнительно ребра модели нагружаются весовым коэффициентом $D(r)$, характеризующим метрику семантической близости текстовых описаний смежных вершин, полученную с помощью моделей машинного обучения для представления вложений текстов отдельных документов (Doc2Vec).

Структура системы построения и анализа семантической модели дескрипторов безопасности объектов ИС представлена на рис. 3. На рис. 3 обозначены: $БД_1$ – документированная база данных для агрегирования слабоструктурированных данных из внешних источников (MongoDB), $БД_2$ – графоориентированная база данных для хранения модели семантического описания сценариев реализации угроз.

Рассмотрим основные этапы предлагаемой методики:

Этап 1. Семантический анализ агрегированного списка рассылок и бюллетеней. Для $БД_1$ формируется поисковый запрос, включающий описание основных узлов ИС, для построения списка семантически близких сообщений специализированных новостных рассылок. Из сформированного списка извлекаются индексы записей CVE и CWE. Поиск осуществляется как на основе оценки меры косинус-расстояния между вектором вложений, построенным с помощью предобученных моделей Doc2Vec для русского и английского языка, так и на основе встроенного в $БД_1$ MongoDB языка запросов. Процедура предобработки и подготовки текстовых данных основана на построении конвейера

NLP-pipe. Далее для формализации признаков строится нейросетевая модель векторного вложения для текстовых документов Distributed memory (PV-DM) [25] Doc2Vec.

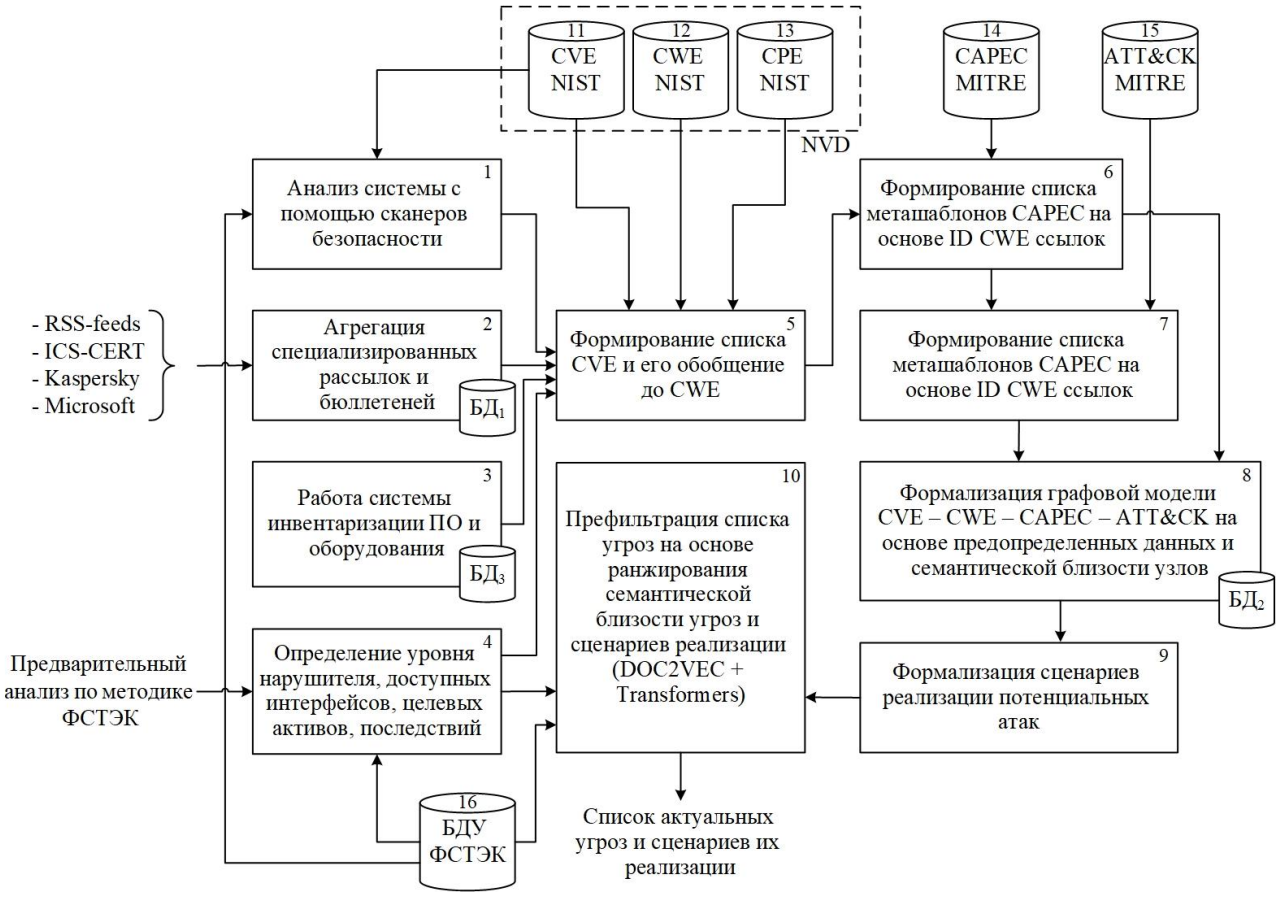


Рис. 3. Структура системы построения и анализа семантической модели дескрипторов безопасности объектов ИС

Этап 2. Анализ результатов работы сканеров безопасности. Формируется список выявленных уязвимостей для каждого компонента ИС в виде перечня CVE-ID и BDU-ID – идентификаторов выявленных уязвимостей.

Этап 3. Фильтрация списка выявленных CVE на основе вектора доступа, сложности атаки, определенных в векторе базовой метрики CVSS для каждой уязвимости, и уровня нарушителя.

Этап 4. Обобщение списка дескрипторов CVE до дескрипторов CWE.

Этап 5. Последовательное выполнение запросов к локальным БД для формирования: списка меташаблонов CAPEC на основе списка CWE-ID, списка техник и тактик ATT&CK на основе списка ID CAPEC.

Этап 6. Формализация графовой модели (на основе задания графа в виде списков вершин и ребер) в $БД_2$ Neo4j с расстановкой весовых коэффициентов на основе оценки меры семантической близости текстовых описаний с поддержкой реализации запросов на языке GraphQL.

Этап 7. Обрезка графовой модели для удаления недостижимых вершин-листьев с повторной оценкой экспертом полученных техник, тактик и сценариев реализации атак (уточнение весового коэффициента и наличия ребра).

Этап 8. Сопоставление текстовых описаний сценариев с базой угроз БДУ ФСТЭК с помощью модели Doc2Vec.

Этап 9. Формализация семантической модели в виде иерархической нечеткой серой когнитивной карты (НСКК), позволяющей анализировать сценарии реализации атак с требуемым уровнем детализации за счет механизмов декомпозиции и укрупнения [23, 24], предложенной в [26].

Исходными данными для конструирования сценариев реализации атаки являются результаты работы экспертов по выявлению уязвимостей элементов ИС, а также потенциальных слабостей программного и аппаратного обеспечения. Наборы показателей системы оценки уязвимостей CVSS и базы данных угроз и уязвимостей позволяют формально описать сценарии эксплуатации уязвимостей и автоматизировать построение цепочки возможных переходов между промежуточными узлами ИС.

В [26] рассматривается процедура «сворачивания» детализированной НСКК, раскрывающей содержание сценариев атак, до укрупненной НСКК уровня представления кибератаки. Каждая атака укрупняется до концепта НСКК с соответствующими весовыми коэффициентами, позволяющими оценить вероятность реализации атаки в каждом из возможных сценариев. Результирующая НСКК позволяет оценить уровень локальных относительных рисков при реализации воздействия нарушителя на промышленную систему. Наиболее детализированный уровень НСКК отражает ряд действий нарушителя на каждом этапе реализации угрозы, что позволяет получить развернутую итоговую оценку локального относительного риска нарушения кибербезопасности для целевых объектов ИС.

Под локальным относительным риском R_i понимается потенциальный ущерб, наносимый *i*-ому активу АСУ ТП предприятия (в относительных единицах) и приводящий к определенным согласно методике ФСТЭК киберфизическим последствиям. Предполагается, что значение риска определяется как установившееся значение состояния *i*-го целевого концепта.

Результирующая НСКК [27] оценки рисков нарушения кибербезопасности для целевых объектов ИС на основе выделенных сценариев реализации атак определяется в виде ориентированного графа, заданного с помощью кортежа множеств:

$$НСКК = \{C, E, W\},$$

где $C = \{C_i\}$ – множество концептов (вершин графа), $(i = 1, 2, ..., n)$; $E = \{E_{ij}\}$ – множество связей между концептами (дуг графа); $W = \{W_{ij}\}$ – множество весов связей, $(i, j) \in \Omega$. Здесь $\Omega = \{(i_1, j_1), (i_2, j_2), ..., (i_S, j_S)\}$ – множество пар индексов смежных (т.е. связанных между собой) вершин графа, $S \le n(n-1)$.

Веса связей НСКК задаются с помощью «серых» (интервальных) чисел $\otimes W_{ij}$, определяемых как $\otimes W_{ij} \in \left[\underline{W}_{ij}, \overline{W}_{ij}\right]$, где $\underline{W}_{ij} < \overline{W}_{ij}$, $\left\{\underline{W}_{ij}, \overline{W}_{ij}\right\} \in [-1, 1]$, где $\underline{W}_{ij}$ и $\overline{W}_{ij}$ – соответственно нижняя и верхняя граница серого числа $\otimes W_{ij}$. Таким

образом, вес связи между *i*-м и *j*-м концептами $\left(C_i \to C_j\right)$ может принимать любое значение в пределах заданного диапазона $\left[\underline{W}_{ij}, \overline{W}_{ij}\right] \in [-1,1]$.

При выборе серых значений весов $\otimes W_{ij}$ экспертам необходимо ориентироваться на нечеткую шкалу (таблице 1).

Таблица 1 – Оценка силы (весов) связей между концептами

Лингвистическое значение силы связи	Обозначение терма	Значение терма	Числовой диапазон
Не влияет	Z	Zero	0
Очень слабая	VS	Very Small	(0;0,15]
Слабая	S	Small	(0,15;0,35]
Средняя	M	Middle	(0,35;0,6]
Сильная	L	Large	(0,6;0,85]
Очень сильная	VL	Very Large	(0,85;1]

Применение методики оценки актуальных угроз и уязвимостей ПО АСУ ТП с использованием методов семантического анализа дескрипторов безопасности и когнитивного моделирования

Рассмотрим пример использования методики анализа уязвимостей и актуальных угроз безопасности информации для фрагмента территориально-распределенной системы обустройства месторождения и транспорта товарной нефти, которая включает основные элементы: добыча нефти, сбор нефти, подготовка нефти, транспортировка товарной нефти.

Ввиду сложности анализируемого объекта, рассмотрим фрагмент базовой модели территориально распределенной системы – АСУ ТП пункта сдачи приема (ПСП) нефти (рис. 4).

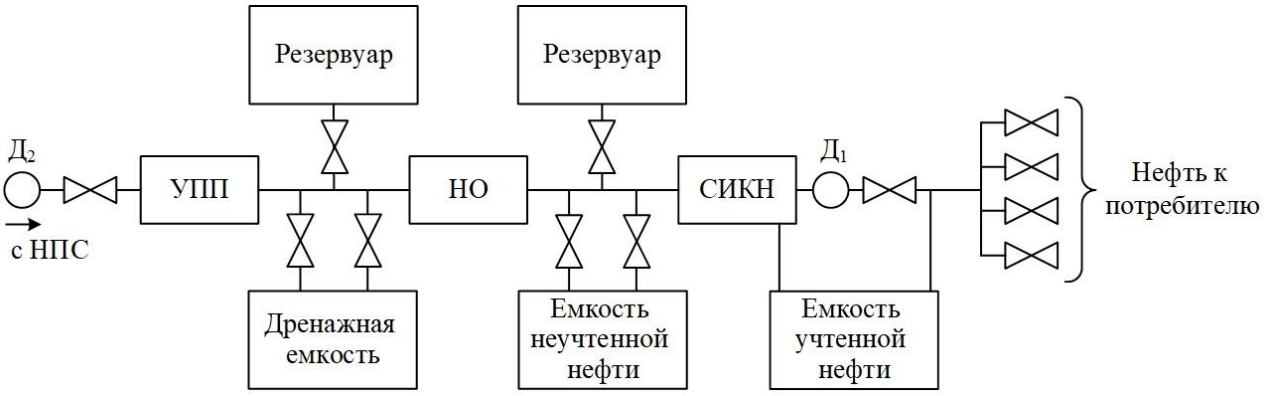


Рис. 4. Базовая модель АСУ ТП ПСП ($Д_1$ – датчики СИКН; $Д_2$ – датчики на входе НПС; УПП – установка подогрева продукта; НО – насосное оборудование; НПС – нефтеперекачивающая станция)

АСУ ТП ПСП в системе магистральных трубопроводов предназначена для автоматизации управления и оперативного контроля технологического процесса,

№3. 2021
ISSN 2410-9916

включая сбор данных о технологических параметрах процесса: расход, уровень, температура, давление, плотность и влажность перекачиваемой нефти.

Подсистемы АСУ ТП ПСП согласно терминологии ГОСТ 62443 можно рассматривать как отдельные зоны безопасности, объединяемые по общим показателям риска, функциональным и/или техническим характеристикам, логическим или физическим границам, сетям передачи данных и т. д. На рис. 5 представлено зонирование по принципу единства выполняемых функций и требований к безопасности их реализации:

- Зона 1 – зона сервера СДКУ (система диспетчерского контроля и управления) SCADA;
- Зона 2 – зона критических устройств управления;
- Зона 3 – зона управления задвижками;
- Зона 4 – зона управления ТП ПСП;
- Зона 5 – зона управления системой измерения количества нефти (СИКН);
- Зона 6 – зона датчиков.

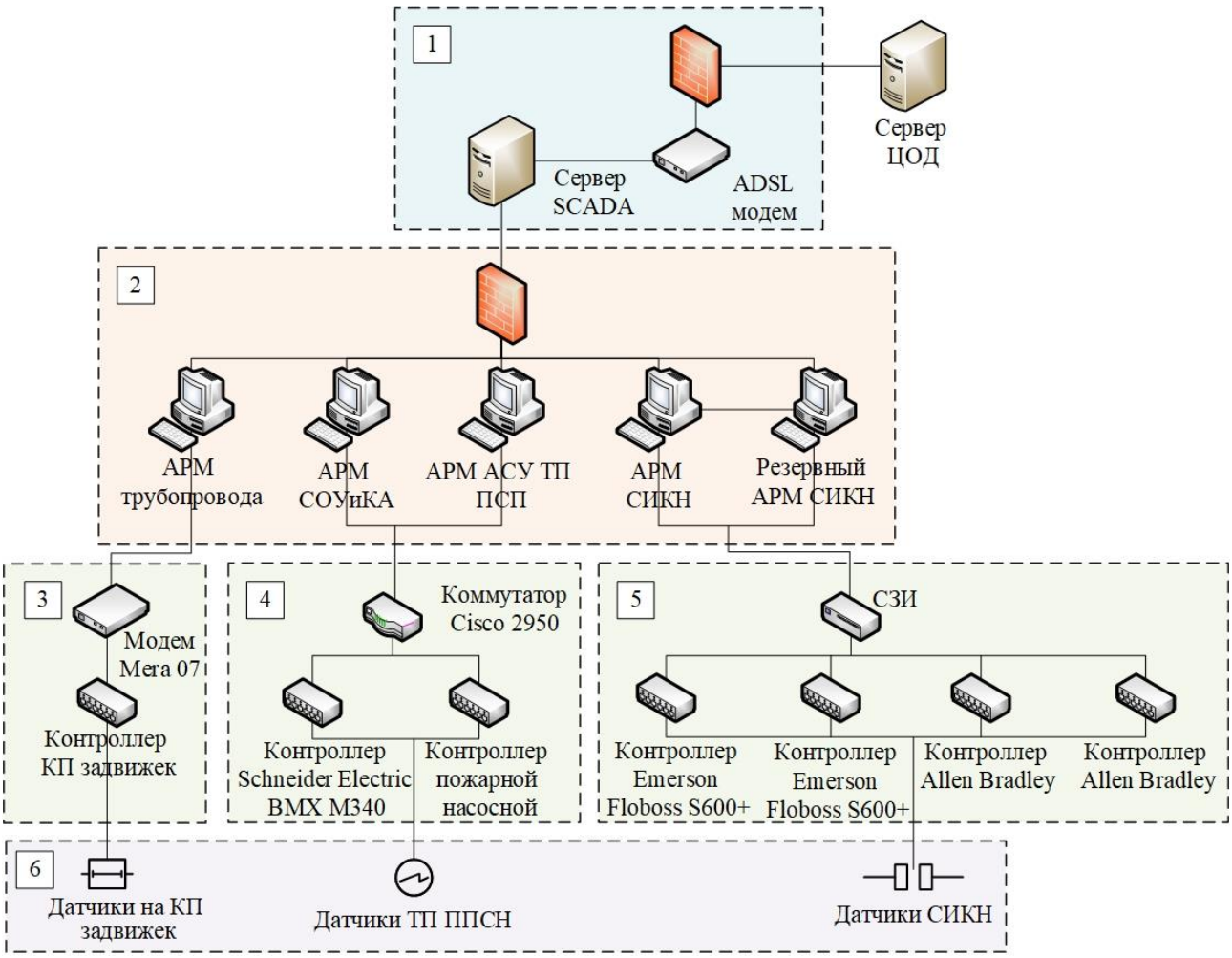


Рис. 5. Зональная модель объекта защиты

Выполним этапы А1-А4 (рис. 1) методики оценки актуальных угроз ФСТЭК для Зоны 5 [10].

DOI: 10.24412/2410-9916-2021-3-110-134
URL: http://sccs.intelgr.com/archive/2021-03/04-Vasilyev.pdf

В Зоне 5 с датчиков СИКН, установленных на двух трубопроводах, данные поступают на два контроллера Emerson Floboss S600+. С двух других датчиков, установленных на третьем и четвертом трубопроводе СИКН, данные поступают на контроллеры Allen Bradley 5561. Два АРМ (основной и резервный) хранят и отображают на мнемосхеме состояние и показатели четырех СИКН в составе ПСП. Дополнительно в операторной установлен АРМ с данными по СИКН для принимающей стороны.

Промежуточные результаты этапов А1-А4 (рис. 1) анализа представлены в таблице 2.

Таблица 2 – Промежуточные результаты реализации методики

Этап		Определяемые параметры	Значения параметров
Определение негативных последствий		Виды рисков (ущерба) и типовые негативные последствия от реализации угроз безопасности информации	– нарушение штатного режима функционирования АСУ ТП (У2) – неспособность выполнять договорные обязательства (У2)
Определение объектов воздействия		Объекты воздействия	Программируемый логический контроллер (ПЛК) для управления насосными станциями
		Виды воздействия	Несанкционированная модификация (изменение) логики работы или уставок ПЛК, которая приводит к включению (или не отключению) насосной станции при закрытой аварийной задвижке в нефтепроводе (У3)
Оценка возможности реализации угроз и их актуальности	Определение цели реализации угроз безопасности информации нарушителями	Виды нарушителя	Авторизованные пользователи систем и сетей
		Категории нарушителя	Внутренний
		Возможные цели реализации угроз безопасности информации	Любопытство или желание самореализации Непреднамеренные, неосторожные или неквалифицированные действия
	Определение уровня возможностей нарушителей по реализации угроз безопасности информации	Уровень возможностей нарушителей	Нарушитель, обладающий базовыми возможностями
		Возможности нарушителей по реализации угроз безопасности информации	нарушители с базовыми возможностями имеют возможность реализовывать только известные угрозы, направленные на известные (документированные) уязвимости, с использованием общедоступных инструментов
	Определения актуальных способов реализации угроз безопасности информации и соответствующие им виды нарушителей и их возможности	Доступные интерфейсы	Локальная вычислительная сеть Веб-интерфейс системы

Для детального моделирования сценариев эксплуатации уязвимостей и определения актуальных угроз воспользуемся предложенной методикой построения семантической модели дескрипторов безопасности объектов АСУ ТП ПСП (рис. 3), последовательно реализуем описанные ранее этапы предложенной методики:

Этап 1. Семантический анализ агрегированного списка рассылок и бюллетеней. Поисковый запрос для $БД_1$ включает описание основных узлов Зоны 5 («промышленная сеть», «промышленный коммутатор», «программируемый логический контроллер»). Процедура предобработки и подготовки текстовых данных извлеченных из CVE, CPE и CWE, представлена в виде конвейера NLP-pipe (таблица 3).

Параметры нейросетевой модели векторного вложения для текстовых документов Distributed memory (PV-DM) Doc2Vec приведены в таблице 4.

Таблица 3 – Структура конвейера предобработки текстовых данных – дескрипторов безопасности объектов АСУ ТП ПСП

Этап		Методы и инструменты
Предобработка	Токенизация	Токенизация на основе адаптивной модели на основе системы правил Razdel [28] (русский язык) и Gensim Simple Tokenizer (английский язык)
	Символьная фильтрация	Регулярные выражения формата regexp, 38 регулярных выражений для фильтрации URLs, HTML-tags и пр.
	Фильтрация токенов	NLTK словари русского и английского языка, дополненные выделенным вручную списком стоп-слов, 400 токенов
Нормализация	Лемматизация	Natasha (русский язык), spacy (английский язык)
Модель вложений	Doc2Vec Distributed Memory	Модель Gensim (минимальная частота встречаемости слова – 2, количество отбрасываемых слов – 5, размер вектора вложений – 100)
	Word2Vec	Модель Gensim (минимальная частота встречаемости слова – 2, количество отбрасываемых слов – 5, размер вектора вложений – 100)
Вектор признаков дескриптора	Композиция	Агрегированный вектор: Doc2Vec и нормированный TF-IDF Word2Vec для лексем в тексте

Таблица 4 – Параметры Distributed memory (PV-DM) Doc2Vec Model

Параметр	Значение
Размерность вектора признаков	100
Размер окна анализа	5
Минимальная частота встречаемости слова для включения в модель	2
Количество эпох обучения	100

Этап 2-4. Формируется список выявленных сканерами безопасности уязвимостей для каждого компонента Зоны 5 в виде перечня CVE-ID и BDU-ID, выполняется его фильтрация и обобщение до дескрипторов CWE.

Этап 5. Список меташаблонов CAPEC, построенный на основе: списка CWE-ID, списка техник и тактик ATT&CK в виде графа приведен на рис. 6.

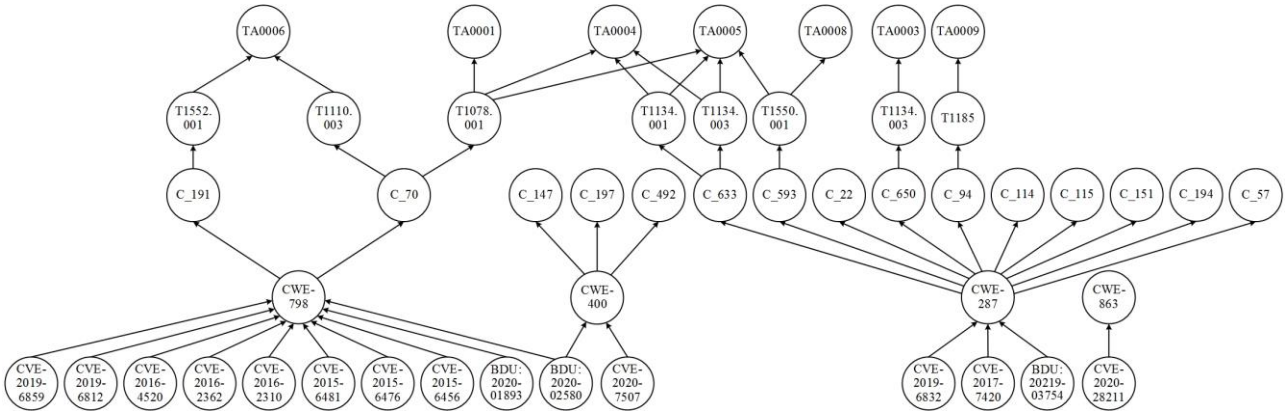


Рис. 6. Фрагмент графа с привязкой CVE-CWE-CAPEC-ATT&CK

Этап 6-7. Формализация графовой модели в БД$_2$ Neo4j с расстановкой весовых коэффициентов на основе оценки меры семантической близости текстовых описаний с поддержкой реализации запросов на языке GraphQL (таблица 5).

Таблица 5 – Матрица связности узлов графовой модели CVE-CWE-CAPEC-ATT&CK по результатам анализа ссылочной модели и семантической близости дескрипторов безопасности объектов

CWE ID	Название	CVE из NVD	CVE из BDU	CAPEC	ATT&CK Technic	ATT&CK Tactics
CWE-798	Use of Hard-coded Credentials (Использование жестко заданных учетных данных)	CVE-2019-6859, CVE-2019-6812, CVE-2016-4520, CVE-2016-2362, CVE-2016-2310, CVE-2015-6481, CVE-2015-6476, CVE-2015-6456.	BDU:2020-01893, BDU:2020-02580	CAPEC-191: Read Sensitive Constants Within an Executable	T 1552.001 Unsecured Credentials:Credentials in files	TA0006 Credential Access (Учетный доступ)
				CAPEC-70: Try Common or Default Usernames and Passwords	T 1078.001 Valid Accounts: Default Accounts (Действительные учетные записи: учетные записи по умолчанию)	TA0005 Defense Evasion, Persistence (Уклонение от защиты, настойчивость); TA0004 Privilege Escalation (Повышение привилегий); TA0001 Initial

CWE ID	Название	CVE из NVD	CVE из BDU	CAPEC	ATT&CK Technic	ATT&CK Tactics
						Access (Первоначальный доступ)
					T 1110.003 Brute Force: Password Spraying (Грубая сила: Распыление паролей)	TA0006 Credential Access (Учетный доступ)
CWE -287	Improper Authentication (Неправильная аутентификация)	CVE-2019-6832, CVE-2017-7420.	BDU:2019 -03754	CAPEC-114, CAPEC-115, CAPEC-22, CAPEC-194, CAPEC-151		
				CAPEC-94	T1185 Man in the Browser (Человек в браузере)	TA0009 Collection (Сбор)
				CAPEC-593	T1550.001 Use Alternate Authentication Material:Application Access Token (Используйте альтернативный материал для аутентификации: токен доступа к приложению)	TA0005 Defense Evasion (Уклонение от защиты); TA0008 Lateral Movement (Боковое движение)
				CAPEC-633	T1134.001 Access Token Manipulation: Token Impersonation/ Theft (Манипулирование токеном доступа: выдача себя за другое лицо / кража токена)	TA0005 Defense Evasion (Уклонение от защиты); TA0004 Privilege Escalation (Повышение привилегий)
					T1134.003 Access Token Manipulation: Make and Impersonate Token (Манипуляции с токенами доступа: создание и выдача токена)	TA0005 Defense Evasion (Уклонение от защиты); TA0004 Privilege Escalation (Повышение привилегий)
				CAPEC-650	T1505.003 Server Software Component: Web Shell (Компонент серверного программного обеспечения: веб-оболочка)	TA0003 Persistence (Настойчивость)
				CAPEC-57		
CWE -863	Incorrect Authorization (Неправильная авторизация)	CVE-2020-28211				
CWE -400	Uncontrolled Resource Consumption (Неконтролируемое потребление ресурсов)	CVE-2020-7507	BDU:2020 -02580	CAPEC-492, CAPEC-197, CAPEC-147		

Этап 8-9. С учетом построенной графовой модели сценариев реализации угроз (тактический уровень моделирования угроз – таблица 5) и результатов анализа согласно методике ФСТЭК (стратегический уровень моделирования угроз – таблица 2), строятся актуальные способы эксплуатации уязвимостей и реализации угроз для элементов Зоны 5 АСУ ТП ПСП в виде НСКК. Итоговая цепочка действий нарушителя в виде укрупненной графовой модели (рис. 7) отражает возможные сценарии эксплуатации уязвимостей для реализации угроз безопасности информации в Зоне 5 АСУ ТП ПСП и позволяет получить оценку локального относительного риска нарушения кибербезопасности целевых концептов промышленной системы. Концепты для моделирования НСКК приведены в таблице 6. Для оценки локального относительного риска используются определенные экспертами значения весов связей между концептами из таблицы 7.

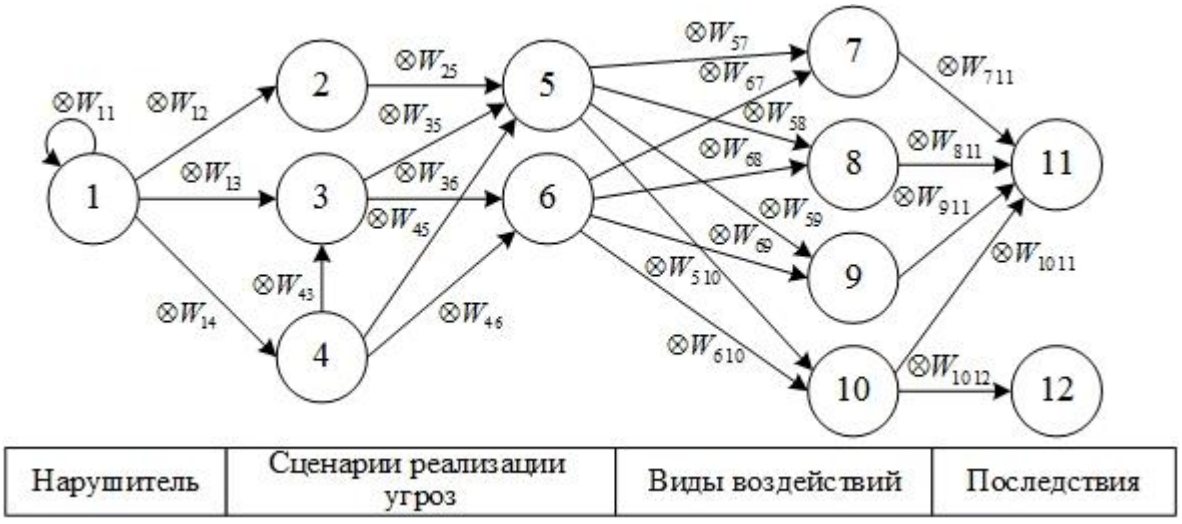


Рис. 7. Укрупненная графовая модель сценариев реализации угроз безопасности информации в Зоне 5 в виде НСКК

Таблица 6 – Концепты НСКК для моделирования актуальных способов реализации угроз

Концепт графовой модели	Характеристика	Группа концептов
C_1	Внутренний нарушитель с базовым потенциалом	
C_2	Подмена ответа сервера FTP резервного копирования конфигураций ПЛК (УБИ.034)	Реализация атак через эксплуатацию недостатков сетевых протоколов
C_3	Перехват учетной записи привилегированного пользователя на ПЛК (УБИ.034)	
C_4	Учетная запись с параметрами по умолчанию на ПЛК (УБИ0.30)	Реализация атак через эксплуатацию недостатков конфигурации или ПО ПЛК
C_5	Модификация прошивки ПЛК (УБИ.188)	Реализация атак через воздействие на управляющую программу ПЛК
C_6	Перезапись проекта ПЛК в режиме online (УБИ.179)	
C_7	Отказ в обслуживании оборудования	Виды воздействия

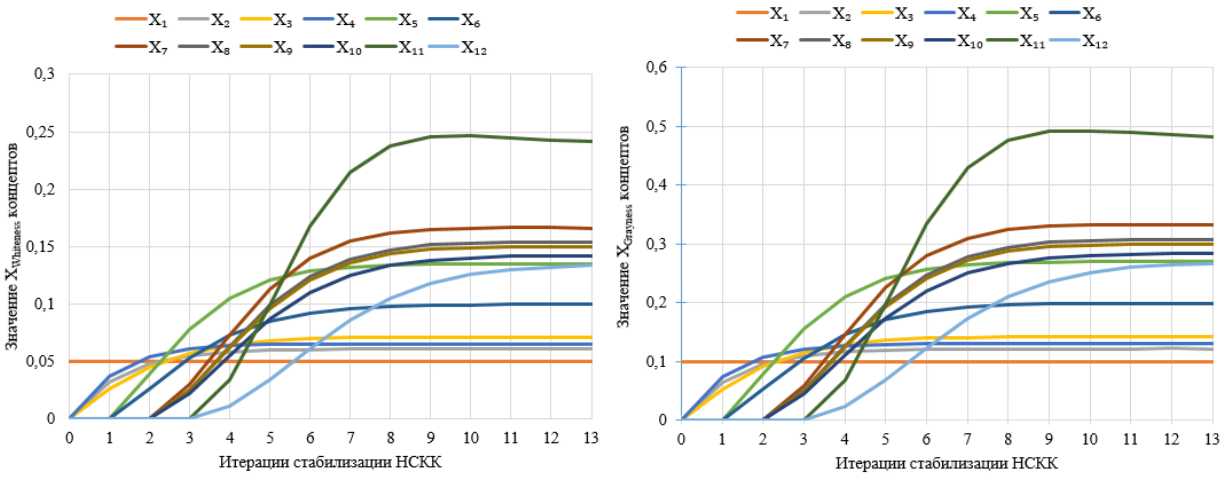
Концепт графовой модели	Характеристика	Группа концептов
C_8	Потеря возможности мониторинга параметров СИКН	
C_9	Перевод СИКН и управляемых объектов в аварийное состояние	
C_{10}	Останов нефтетранспорта по магистральному нефтепроводу	
C_{11}	Нарушение штатного режима функционирования АСУ ТП ПСП	Последствия
C_{12}	Неспособность компании выполнить договорные обязательства	

Таблица 7 – Веса связей НСКК оценки рисков реализации угроз безопасности информации для целевых концептов ИС

Вес связи	Диапазон	Вес связи	Диапазон
W_{11}	[0,4; 0,45]	W_{58}	[0,6; 0,85]
W_{12}	[0,3; 0,5]	W_{59}	[0,5; 0,65]
W_{13}	[0,3; 0,45]	$W_{5\,10}$	[0,3; 0,45]
W_{14}	[0,5; 0,7]	W_{67}	[0,6; 0,8]
W_{25}	[0,5; 0,7]	W_{68}	[0,25; 0,4]
W_{26}	[0,25; 0,4]	W_{69}	[0,5; 0,7]
W_{35}	[0,5; 0,75]	$W_{6\,10}$	[0,5; 0,75]
W_{36}	[0,5; 0,6]	$W_{7\,11}$	[0,35; 0,55]
W_{43}	[0,15; 0,25]	$W_{8\,11}$	[0,25; 0,45]
W_{45}	[0,25; 0,4]	$W_{9\,11}$	[0,5; 0,6]
W_{46}	[0,2; 0,3]	$W_{10\,11}$	[0,65; 0,85]
W_{57}	[0,6; 0,8]	$W_{10\,12}$	[0,7; 0,85]

Концепт C_1 выступает в качестве концепта-драйвера и представляет собой внутреннего нарушителя с базовым потенциалом при реализации угрозы, связанной с модификацией конфигураций ПЛК с целью нарушения технологического процесса, создания аварийной ситуации на промышленном объекте или состояния аварийной остановки.

Процесс изменения состояний концептов НСКК во времени показан на рис. 8, где по оси ординат расположены значения переменных состояния НСКК, а по оси абсцисс – итерации сходимости НСКК.



а) б)

Рис. 8. Изменение во времени состояний концептов НСКК: (а) стабилизация «белого» значения концепта (б) стабилизация «серости» концепта

На графиках приведены центральные значения «серых» переменных состояния концептов (рис. 8, а) и значения «серости» этих концептов (рис. 8, б). Из графиков видно, что состояния концептов стабилизировались за 8-10 итераций. Значения концептов НСКК, соответствующие оценкам локальных относительных рисков нарушения штатного режима функционирования АСУ ТП ПСП, составило [0,2473; 0,7231], неспособности компании выполнить договорные обязательства – [0,0851; 0,3538]. Усредненные значения локальных относительных рисков в рассматриваемых сценариях реализации угроз составили $X_{11}^{*} = 0,4852$ и $X_{12}^{*} = 0,2195$ соответственно.

Заключение

Предлагаемая методика основан на построении семантической модели дескрипторов безопасности с целью автоматизации низкоуровневого моделирования сценариев реализации угроз на основе описания шаблонов компьютерных атак (CAPEC, ATT&CK, OWASP, STIX, WASC и др.). Отличительной особенностью семантической модели является использование и ссылочной модели, и оценки семантической близости дескрипторов из отечественных и зарубежных баз данных. Это позволяет снизить трудоемкость формирования перечня актуальных угроз и уязвимостей за счет применения технологий Text Mining для префильтрации несвязанных или недостижимых вершин при выполнении основных этапов анализа согласно методике ФСТЭК России. Завершающий этап предлагаемой методики позволяет перейти к построению когнитивной модели оценки рисков нарушения кибербезопасности для целевых объектов АСУ ТП, что позволяет получить детализированную оценку рисков и сделать более обоснованный выбор средств защиты информации за счет возможности моделирования различных сценариев реализации угрозы. Исходными данными для построения когнитивных карт являются не только экспертные оценки, но и формализованные и систематизированные данные из открытых баз данных

угроз и уязвимостей, что существенно повышает обоснованность и полноту моделирования.

Автоматизированное моделирование и оценка актуальности угроз и сценариев их реализации на основе перечня выявленных уязвимостей для всех компонентов АСУ ТП позволяет выявить наиболее вероятные сценарии реализации угроз и оценить последствия от их реализации.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-08-00668 (когнитивное моделирование) и № 20-38-90078 (семантическая модель дескрипторов).

Литература

1. Актуальные киберугрозы: IV квартал 2020 года // Отчет Positive Technologies [Электронный ресурс]. – URL: https://www.ptsecurity.com/ru-ru/research/analytics/cybersecurity-threatscape-2020-q4/ (дата обращения 08.04.2021).

2. Ландшафт угроз для систем промышленной автоматизации. 2019 год // Kaspersky ICS CERT [Электронный ресурс]. – URL: http:// www.cert.kaspersky.ru (дата обращения 08.04.2021).

3. What is Cyber Threat Intelligence? // Cisco Systems, Inc. [Электронный ресурс]. – URL: https://www.cisco.com/c/en/us/products/security/what-is-cyber-threat-intelligence.html (дата обращения 08.04.2021).

4. Котерева О. Откуда возникла необходимость в Threat Intelligence // Информзащита [Электронный ресурс]. – URL: https://habr.com/ru/company/infosecurity/blog/336676/ (дата обращения 08.04.2021).

5. Smyth V. Vulnerability Intelligence // ITNOW. 2016. P. 26-27. doi: 10.1093/itnow/bww100.

6. Котенко И. В., Дойникова Е. В., Чечулин А. А. Общее перечисление и классификация шаблонов атак (CAPEC): описание и примеры применения // Защита информации. Инсайд. 2012. № 4 (46). С. 54-66.

7. Федорченко А. В., Чечулин А. А., Котенко И. В. Исследование открытых баз уязвимостей и оценка возможности их применения в системах анализа возможности их применения в системах анализа защищенности компьютерных сетей // Информационно-управляющие системы. 2014. № 5. С. 72-79.

8. Калинин Н., Шерварлы В., Петухов А. Общий обзор классификаций угроз безопасности: OWAPS, CWE, CAPEC, WASC [Электронный ресурс]. – URL: https://safe-surf.ru/specialists/article/5210/595970/ (дата обращения 08.04.2021).

9. Сапожников А. Общий обзор реестров и классификаций уязвимостей (CVE, OSVDB, NDV, Secunia) [Электронный ресурс]. – URL: https://safe-surf.ru/specialists/article/5228/607311/ (дата обращения 08.04.2021).

10. Методика оценки угроз безопасности информации. Методический документ ФСТЭК России от 5 февраля 2021 г. // Официальный сайт ФСТЭК

России [Электронный ресурс]. – URL: https://fstec.ru/component/attachments/download/2919 (дата обращения 08.04.2021).

11. Бондарчук Д. В. Векторная модель представления знаний на основе семантической близости термов // Вестник Южно-Уральского государственного университета. Серия: Вычислительная математика и информатика. 2017. Т. 6. № 3. С. 73-83. doi: 10.14529/cmse170305.

12. Word2Vec: как работать с векторными представлениями слов // Neurohive (Базовый курс). [Электронный ресурс]. – URL: https://neurohive.io/ru/osnovy-data-science/word2vec-vektornye-predstavlenija-slov-dlja-mashinnogo-obuchenija/ (дата обращения 08.04.2021).

13. Бенгфорт Б., Билбро Р., Океда Т. Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка / Пер. с англ. – СПб.: Питер, 2019. – 368 с.

14. Lee Y., Shin S. Toward Semantic Assessment of Vulnerability Severity: A Text Mining Approach // Proceedings of ACM CIKM Workshop (EYRE'18). – URL: https://www.CEUR-WS.org/Vol1-2482/papers.pdf (дата обращения 08.04.2021).

15. Noel S. Text Mining for Modeling Cyberattacks // Chapter 14 in the book: Handbook of Statistics. Elsevier B.V. (Part C: Applications and Linguistic Diversity). 2018. Vol. 38. P. 461-515. doi: 10.1016 / bs.host.2018.06.001.

16. Hemberg E. et al. Linking Threat Tactics, Techniques, and Patterns with Defensive Weaknesses, Vulnerabilities and Affected Platform Configurations for Cyber Hunting // arXiv e-prints [Электронный ресурс]. – URL: https://arxiv.org/pdf/2010.00533v2.pdf (дата обращения 08.04.2021).

17. Xiao H. et al. Embedding and Predicting Software Security Entity Relationships: A Knowledge Graph Based Approach // International Conference on Neural Information Processing. Springer, Cham, 2019. P. 50-63.

18. Noel S. Interactive visualization and text mining for the CAPEC cyber attack catalog // Proceedings of the ACM Intelligent User Interfaces Workshop on Visual Text Analytics. [Электронный ресурс]. – URL: https://csis.gmu.edu/noel/pubs/2015_CAPEC_viz.pdf (дата обращения 08.04.2021).

19. Wareus E., Martin H. Automated CPE Labeling of CVE Summaries with Machine Learning. International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment // 17th International Conference. Lisbon, 2020. Vol. 12223. P. 3-22. doi:10.1007/978-3-030-52683-2_1.

20. Datta P., Lodinger N., Namin S., Jones S. Cyber-Attack Consequence Prediction // Proceedings of the 3rd Workshop on Big Data Engineering and Analytics in Cyber-Physical Systems. 9 p. [Электронный ресурс]. – URL: https://arxiv.org/abs/2012.00648v2 (дата обращения 08.04.2021).

21. Лукацкий А. В. Как я моделировал угрозы по проекту новой методики ФСТЭК. Часть 1 [Электронный ресурс]. – URL: https://lukatsky.blogspot.com/2020/04/1.html (дата обращения 08.04.2021).

22. Васильев В. И., Вульфин А. М., Кучкарова Н. В. Автоматизация анализа уязвимостей программного обеспечения на основе технологии Text Mining // Вопросы кибербезопасности. 2020. № 4 (38). С. 22-31. doi:10.21681/2311-3456-2020-04-22-31.

23. Васильев В. И., Вульфин А. М., Герасимова И. Б., Картак В. М. Анализ рисков кибербезопасности с помощью нечетких когнитивных карт // Вопросы кибербезопасности. 2020. № 2 (36). С. 11-21. doi:10.21681/2311-3456-2020-2-11-21.

24. Васильев В. И., Вульфин А. М., Гузаиров М. Б., Картак В. М., Черняховская Л. Р. Оценка рисков кибербезопасности АСУ ТП промышленных объектов на основе вложенных нечетких когнитивных карт // Информационные технологии. 2020. Т. 26. № 4. С. 213–221. doi: 10.17587/it.26.213-221.

25. Mendsaikhan O. et al. Identification of cybersecurity specific content using the Doc2Vec language model // 2019 IEEE 43rd annual computer software and applications conference (COMPSAC). IEEE, 2019. Vol. 1. P. 396-401.

26. Васильев В. И., Кириллова А. Д., Вульфин А. М. Когнитивное моделирование вектора кибератак на основе меташаблонов CAPEC // Вопросы кибербезопасности, 2021. № 2 (42). С. 2-16. DOI: 10.21681/2311-3456-2021-2-2-16.

27. Zhang J. Y., Liu Z. Q., Zhou S. Quotient FCMs-a decomposition theory for fuzzy cognitive maps // IEEE transactions on fuzzy systems. 2003. Vol. 11 (5). P. 593-604. doi: 10.1109/TFUZZ.2003.817836.

28. Rule-based token, sentence segmentation for Russian language [Электронный ресурс]. – URL: https://github.com/natasha/razdel (дата обращения 08.04.2021).

References

1. Aktual'nye kiberugrozy: IV kvartal 2020 goda [Topical Cyber Threats: Q4 2020]. *Positive Technologies*. Available at: https://www.ptsecurity.com/ru-ru/research/analytics/cybersecurity-threatscape-2020-q4/ (accessed 8 April 2021) (in Russian).

2. Landshaft ugroz dlia sistem promyshlennoi avtomatizatsii. 2019 god [Threat landscape for industrial automation systems. 2019]. *Kaspersky ICS CERT*. Available at: http://www.cert.kaspersky.ru (accessed 8 April 2021) (in Russian).

3. What is Cyber Threat Intelligence? *Cisco Systems, Inc.* Available at: https://www.cisco.com/c/en/us/products/security/what-is-cyber-threat-intelligence.html (accessed 8 April 2021).

4. Kotereva O. Otkuda voznikla neobkhodimost' v Threat Intelligence [Where did the need for Threat Intelligence come from?]. *Informzashchita*. Available at: https://habr.com/ru/company/infosecurity/blog/336676/ (accessed 8 April 2021) (in Russian).

5. Smyth V. Vulnerability Intelligence. *ITNOW*, 2016, pp. 26-27. doi: 10.1093/itnow/bww100.

6. Kotenko I. V., Doinikova E. V., Chechulin A. A. Obshchee perechislenie i klassifikatsiia shablonov atak (CAPEC): opisanie i primery primeneniia [General Enumeration and Classification of Attack Patterns (CAPEC): Description and Application Examples]. *Zaŝita informacii. Inside*, 2012, no. 4(46), pp. 54-66 (in Russian).

7. Fedorchenko A. V., Chechulin A. A., Kotenko I. V. Issledovanie otkrytyh baz uyazvimostej i ocenka vozmozhnosti ih primeneniya v sistemah analiza

vozmozhnosti ih primeneniya v sistemah analiza zashchishchennosti komp'yuternyh setej [Open Vulnerability Bases and their Application in Security Analysis Systems of Computer Networks]. *Informatsionno-upravliaiushchie sistemy*, 2014, no. 5, pp. 72-79 (in Russian).

8. Kalinin N., Shervarly V., Petukhov A. Obshchii obzor klassifikatsii ugroz bezopasnosti: OWAPS, CWE, CAPEC, WAS [An overview of the classifications of security threats: OWAPS, CWE, CAPEC, WAS]. Available at: https://safe-surf.ru/specialists/article/5210/595970/ (accessed 8 April 2021) (in Russian).

9. Sapozhnikov A. Obshchii obzor reestrov i klassifikatsii uiazvimostei (CVE, OSVDB, NDV, Secunia) [General overview of registries and classifications of vulnerabilities (CVE, OSVDB, NDV, Secunia)]. Available at: https://safe-surf.ru/specialists/article/5228/607311/ (accessed 8 April 2021) (in Russian).

10. Methods for assessing threats to information security. Methodological document of the FSTEC of Russia dated February 5, 2021. *Official site of FSTEC of Russia*. Available at: https://fstec.ru/component/attachments/download/2919 (accessed 8 April 2021) (in Russian).

11. Bondarchuk D. V. Vektornaya model' predstavleniya znanij na osnove semanticheskoj blizosti termov [Vector Space Model of Knowledge representation Based on Semantic Relatedness]. *Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering*, 2017, vol. 6, no. 3. pp. 73-83. doi: 10.14529/cmse170305 (in Russian).

12. Word2Vec: kak rabotat' s vektornymi predstavleniiami slov [Word2Vec: how to work with vector representations of words]. *Neurohive*. Available at: https://neurohive.io/ru/osnovy-data-science/word2vec-vektornye-predstavlenija-slov-dlja-mashinnogo-obuchenija/ (accessed 8 April 2021) (in Russian).

13. Bengfort B., Bilbro R., Ojeda T. *Mobile Applied Text Analysis with Python. Enabling Language-Aware Data Products with Machine Learning*. O'Reilly Media, Inc, 2019. 364 p.

14. Lee Y., Shin S. Toward Semantic Assessment of Vulnerability Severity: A Text Mining Approach. *Proceedings of ACM CIKM Workshop (EYRE'18).* 2018. Available at: https://www.CEUR-WS.org/Vol1-2482/papers.pdf accessed 8 April 2021).

15. Noel S. Text Mining for Modeling Cyberattacks. *Chapter 14 in the book: Handbook of Statistics. Elsevier B.V.* (Part C: Applications and Linguistic Diversity), 2018, vol. 38, pp. 461-515. doi: 10.1016 / bs.host.2018.06.001.

16. Hemberg E. et al. Linking Threat Tactics, Techniques, and Patterns with Defensive Weaknesses, Vulnerabilities and Affected Platform Configurations for Cyber Hunting. *DeepAI: The front page of A.I.*, 2020, p. 14. Available at: https://arxiv.org/pdf/2010.00533v2.pdf (accessed 8 April 2021).

17. Xiao H., Xing Z., Guo H. Embedding and Predicting Software Security Entity Relationships: A Knowledge Graph Based Approach. International Conference on Neural Information Processing. *Springer, Cham*, 2019, pp. 50-63. Available at: https://arxiv.org/pdf/1903.04750v1.pdf (accessed 8 April 2021).

18. Noel S. Interactive visualization and text mining for the CAPEC cyber attack catalog. *Proceedings of the ACM Intelligent User Interfaces Workshop on*

Visual Text Analytics, 2015. Available at: https://csis.gmu.edu/noel/pubs/2015_CAPEC_viz.pdf (accessed 8 April 2021).

19. Wåreus E., Martin H. Automated CPE Labeling of CVE Summaries with Machine Learning. International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment. *17th International Conference. Lisbon*, 2020, vol. 12223. pp. 3-22. doi:10.1007/978-3-030-52683-2_1.

20. Datta P., Lodinger N., Namin S., Jones S. Cyber-Attack Consequence Prediction. *Proceedings of the 3rd Workshop on Big Data Engineering and Analytics in Cyber-Physical Systems*. 9 p, 2020. Available at: https://arxiv.org/abs/2012.00648v2 (accessed 8 April 2021).

21. Lukackiy A. V. Kak ya modeliroval ugrozy po proektu novoj metodiki FSTEK. Chast' 1 [How I Modeled Threats Based on the New FSTEC Methodology Project. Part 1]. Available at: https://lukatsky.blogspot.com/2020/04/1.html (accessed 8 April 2021).

22. Vasilyev V. I., Vulfin A. M., Kuchkarova N. V. Automation of software vulnerabilities analysis on the basis of Text Mining technology. *Voprosy kiberbezopasnosti*, 2020, no. 4 (38), pp. 22-31. doi:10.21681/2311-3456-2020-04-22-31.

23. Vasilyev V. I., Vulfin A. M., Gerasimova I. B., Kartak V. M. Analysis of cybersecurity risk with use of fuzzy cognitive maps. *Voprosy kiberbezopasnosti*, 2020, no. 2 (36), pp. 11-21. doi:10.21681/2311-3456-2020-2-11-21.

24. Vasilyev V. I., Vulfin A. M., Guzairov M. B., Kartak V. M., Chernjahovskaja L. R. Cybersecurity risk assessment of industrial objects' ACS of TP on the basis of nested fuzzy cognitive maps technology. *Informacionnye tehnologii*, 2020, vol. 4 (26), pp. 213-221.

25. Mendsaikhan O. et al. Identification of cybersecurity specific content using the Doc2Vec language model. *2019 IEEE 43rd annual computer software and applications conference (COMPSAC). IEEE*, 2019, vol. 1, pp. 396-401.

26. Vasilyev V. I., Kirillova A. D., Vulfin A. M. Cognitive modeling of the cyber attack vector based on CAPEC methods. *Voprosy kiberbezopasnosti*, 2021, vol. 2 (42), pp. 2-16. doi: 10.21681/2311-3456-2021-2-2-16.

27. Zhang J. Y., Liu Z. Q., Zhou S. Quotient FCMs-a decomposition theory for fuzzy cognitive maps. *IEEE transactions on fuzzy systems*, 2003, vol. 11 (5), pp. 593-604. doi: 10.1109/TFUZZ.2003.817836.

28. Rule-based token, sentence segmentation for Russian language Available at: https://github.com/natasha/razdel (accessed 8 April 2021).

Статья поступила 27 апреля 2021 г.

Информация об авторах

Васильев Владимир Иванович – доктор технических наук, профессор. Профессор кафедры вычислительной техники и защиты информации. Уфимский государственный авиационный технический университет. Область научных интересов: интеллектуальные системы управления; информационная безопасность. E–mail: vasilyev@ugatu.ac.ru

Вульфин Алексей Михайлович – кандидат технических наук. Доцент кафедры вычислительной техники и защиты информации. Уфимский государственный авиационный технический университет. Область научных интересов: интеллектуальный анализ данных; моделирование сложных технических систем. E–mail: vulfin.alexey@gmail.com

Кириллова Анастасия Дмитриевна – аспирант кафедры вычислительной техники и защиты информации. Уфимский государственный авиационный технический университет. Область научных интересов: обеспечение информационной безопасности; оценка рисков кибербезопасности. E–mail: kirillova.andm@gmail.com

Кучкарова Наиля Вакилевна – старший преподаватель кафедры вычислительной техники и защиты информации. Уфимский государственный авиационный технический университет. Область научных интересов: обеспечение информационной безопасности; семантический анализ текстов. E–mail: nailya_kuchkarov@mail.ru

Адрес: 450008, Россия, г. Уфа, ул. К. Маркса, д. 12.

Methodology for Assessing Current Threats and Vulnerabilities Based on Cognitive Modeling Technologies and Text Mining

V. I. Vasilyev, A. M. Vulfin, A. D. Kirillova, N. V. Kuchkarova

Formulation of the problem. *Automation of low-level modeling of scenarios for exploiting vulnerabilities and implementing threats based on the description of vectors (patterns) of computer attacks contained in databases that characterize various aspects of the security of software and hardware infrastructure of a network of industrial facilities.* ***Purpose.*** *Development of a methodology for automated assessment of current threats and vulnerabilities according to the FSTEC of Russia Methodology.* ***Methods:*** *semantic modeling based on cross-referencing and assessing the degree of similarity of textual descriptions of objects of a network of industrial facilities borrowed from the BDU FSTEC and CAPEC, MITRE and ATT&CK databases. Modeling is based on the formalization of a logical chain: "many identified software vulnerabilities; many relevant threats; many of the most likely attack patterns (threat scenarios); possible cyber-physical consequences; quantitative assessment of cybersecurity risk", taking into account the requirements of the FSTEC of Russia regulations. The formalization is performed in the notation of graph semantic models and fuzzy cognitive models.* ***Novelty.*** *The semantic graph model extends the reference model of the relationship between software and hardware security aspects. The difference of the model lies in the use of weight coefficients based on an estimate of the degree of proximity of the vectors of Doc2Vec nesting of text descriptions of objects. The use of the model makes it possible to reduce the labor intensity of forming a list of current threats and vulnerabilities through the use of Text Mining technologies for pre-filtering unrelated or unattainable peaks when performing the main stages of analysis according to the FSTEC of Russia Methodology. Building a cognitive model for assessing the risks of cybersecurity breach for target ICS objects based on the semantic model allows you to obtain a detailed assessment of cybersecurity risks and make a more informed choice of information protection tools.* ***Results.*** *A methodology for assessing actual (i.e. potentially most dangerous) vulnerabilities and threats to ICS software is presented. An example of the application of the proposed methodology for assessing the threats and vulnerabilities of the applied software of the ICS subsystem of an industrial oil production facility is considered, followed by a quantitative assessment of the risks of cybersecurity breaches.* ***Practical relevance.*** *The proposed approach makes it possible to automate low-level modeling of scenarios for the implementation of threats based on the description of vectors (patterns) of computer attacks and to reduce the laboriousness of developing the formation of a list of actual threats through the use of Text Mining technologies for pre-filtering unrelated or unattainable peaks when performing the main stages of analysis according to the FSTEC of Russia Methodology.*

Key words: *vulnerability; threat; cyberattack; classification systems; semantic analysis of texts.*

Information about Authors

Vladimir Ivanovich Vasilyev – Dr. habil. of Engineering Sciences, Full Professor. Professor at the Department of Computer Engineering and Information Security. Ufa State Aviation Technical University. Field of research: intelligent control systems; information security. E–mail: vasilyev@ugatu.ac.ru

Alexey Mikhailovich Vulfin – Ph.D. of Engineering Sciences. Associate Professor at the Department of Computer Engineering and Information Security. Ufa State Aviation Technical University. Field of research: data mining; modeling of complex technical systems. E–mail: vulfin.alexey@gmail.com

Anastasia Dmitrievna Kirillova – Postgraduate at the Department of Computer Engineering and Information Security. Ufa State Aviation Technical University. Field of research: information security, cybersecurity risk assessment. E–mail: kirillova.andm@gmail.com

Nailya Vakilevna Kuchkarova – Senior Lecturer at the Department of Computer Engineering and Information Security. Ufa State Aviation Technical University. Field of research: information security, semantic analysis of texts. E–mail: nailya_kuchkarov@mail.ru

Address: 450008, Russia, Ufa, Karl Marx Str., 12.