ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ
САНКТ-ПЕТЕРБУРГСКИЙ ИНСТИТУТ ИНФОРМАТИКИ И АВТОМАТИЗАЦИИ
РОССИЙСКОЙ АКАДЕМИИ НАУК
(СПИИРАН)
лаборатория проблем компьютерной безопасности
профильная лаборатория

Федорченко Андрей Владимирович

Корреляция больших массивов гетерогенных данных для мониторинга и управления безопасностью в киберфизических системах

научно – квалификационная работа (диссертация)
по основной образовательной программе подготовки научно-педагогических кадров в аспирантуре

направление подготовки 09.06.01 «Информатика и вычислительная техника»
направленность 05.13.01 «Системный анализ, управление и обработка информации»

Научный руководитель
д-р техн. наук, профессор

_____ _____И.В. Котенко_____
подпись *И.О.Фамилия*

«_____»_____ 2018 г.

Автор работы
аспирант

_____ _____А.В. Федорченко____
подпись *И.О.Фамилия*

Санкт-Петербург – 2018

Оглавление

Введение

Актуальность темы научно-квалификационной работы. По мере развития информатики и техники, в настоящее время все более актуальными становятся *киберфизические системы* (КФС), как объекты синтеза информационных и технологических процессов. Данные системы как правило встречаются в высокотехнологичных промышленных, научных, транспортных и военных отраслях, таких как: ядерная, топливная и гидро- энергетика, авиация и космонавтика, высокоскоростное сообщение, противоракетная оборона, навигация и многие другие. Как правило, КФС в данных отраслях предназначены для обслуживания критически важных объектов (КВО). Выход из строя или нарушение работы КВО могут привести к различным негативным последствиям, таким как: сбои в техногенных комплексах, нарушение экологии и человеческие жертвы. Исходя из важности сохранения отказоустойчивости КФС, задача обеспечения безопасности объектов, в которых они применяются, является достаточно актуальной.

На данный момент существует обширное количество классов средств обеспечения безопасности компьютерных инфраструктур. К данным классам относятся: системы обнаружения (предотвращения) вторжений (Intrusion Detection (Prevention) System, IDS (IPS)), антивирусы, сканеры уязвимостей (сканеры безопасности), межсетевые экраны и другие. Также, одним из классов средств защиты являются системы мониторинга и управления безопасностью, или системы управления информацией и событиями безопасности (Security Information and Event Management, SIEM). Сегодня SIEM-системы активно развиваются и являются одними из наиболее обширных средств контроля защищенности компьютерных инфраструктур [1,2]. Данный факт обусловлен изначально поставленной перед средствами защиты указанного класса задачей: управление информацией и событиями безопасности, предполагая корреляцию разнородных и разноуровневых событий (инцидентов). Иными словами, работа

SIEM-систем направлена, в том числе на интеграцию информации и событий, полученных от различных средств обеспечения безопасности.

Корреляция данных для обеспечения безопасности в киберфизических системах (КФС) является важной и актуальной задачей, поскольку их инфраструктуры отличаются высокой гетерогенностью и объемами информации, а существующие методы не в состоянии эффективно выполнять задачи корреляции в них. К системам данного класса относятся сети «Интернета вещей» (Internet of Things, IoT), автоматизированные системы управления технологическим процессом (АСУ ТП), бортовые компьютеры транспорта с функцией корректировки курса и многие другие.

Основная проблема выполнения корреляции данных в КФС обусловлена структурной, конфигурационной и функциональной неопределенностью их инфраструктуры. Эффективность выполнения задачи корреляции значительно снижается при высокой гетерогенности источников информации, несогласованности их форматов представления данных, условно-неограниченном количестве, а также возрастающей сложности и скрытости потенциальных атак. Одним из классов средств защиты информации в подобных условиях являются системы управления информацией и событиями безопасности (Security Information and Event Management, SIEM) следующего поколения. Предполагается, что для выполнения задач корреляции в неопределенных инфраструктурах необходим этап предварительного анализа данных, в результате которого производится автоматизированная адаптация система корреляции к целевой инфраструктуре, а именно, определение ее структуры, типов активов и их иерархии. Дальнейший анализ направлен на обучение моделей поведения различных объектов целевой инфраструктуры (ЦИ) с целью проактивного мониторинга, позволяющего заранее вычислить вероятность наступления конкретного инцидента.

Различные методы корреляции данных исследуются научным сообществом на протяжении более 20 лет. Однако существующие методики, как правило, требуют тонкой ручной либо автоматизированной настройки под конкретную

задачу, а на практике преимущественно применяются методы на основе правил. Таким образом, разработка адаптивного подхода корреляции данных в условиях неопределенной инфраструктуры является актуальной темой исследований.

Степень разработанности темы научно-квалификационной работы. Вопросам корреляции гетерогенных данных для различных прикладных задач посвящено большое количество работ как отечественных исследователей: О.Ю. Воробьев, С.В. Клочков, И.В. Котенко, И.Б. Саенко, так и зарубежных: C. Kruegel, F. Valeur, G. Vigna, R. Sadoddin, A. Ghorbani, A. Muller, T. Limmer, F. Dressler и др. Анализ предметной области показал, что предлагаемые методики преимущественно не обладают возможностью комплексной адаптации к ЦИ, а существующие подходы адаптации направлены на частные случаи неопределенностей (неопределенность поведения, неопределенность состояния и др.). В связи с данным фактом в текущем исследовании была поставлена задача разработки комплексного подхода к корреляции данных для выполнения задач безопасности в SIEM-системах на основе адаптации к неопределенной инфраструктуре КФС.

Научная задача. Разработка модельно-методического аппарата для выполнения процесса корреляции данных безопасности в КФС на основе адаптации к их инфраструктурам в условиях неопределенности.

Объект исследования. КФС, а также их информация и события безопасности.

Предмет исследования. Модели, методики и алгоритмы корреляции условно-неограниченного количества разнородных событий и информации безопасности в КФС.

Основной целью научно-квалификационной работы является повышение эффективности выполнения процесса корреляции событий и информации безопасности в КФС за счет применения комбинированной методики корреляции, ориентированной на обработку исходных данных от условно-неограниченного количества гетерогенных источников и адаптацию к неопределенной ЦИ. Для достижения указанной цели в работе поставлены и решены следующие **задачи**:

1. Анализ информации и событий безопасности, а также методов и подходов к их корреляции для формального описания исходных данных и формирования модели неопределенной ЦИ КФС.

2. Разработка модели корреляции данных безопасности с условно-статичным содержимым на основе онтологического подхода.

3. Разработка модели и комплексной методики корреляции данных с динамичным содержимым (событий) с возможностью автоматизированной адаптации к неопределенной ЦИ КФС.

4. Построение архитектуры и реализация программного прототипа на основе предложенных моделей и методик.

5. Экспериментальная оценка предложенных алгоритмов и методик по показателям оперативности, масштабируемости и ресурсопотребления.

Положения, выносимые на защиту:

1. Онтологическая модель корреляции данных безопасности с условно-статичным содержимым.

2. Модель корреляции данных безопасности с динамичным содержимым.

3. Комплексная методика корреляции информации и событий безопасности с адаптацией к неопределенной ЦИ КФС.

Научная новизна диссертационной работы состоит в следующем:

1. Предложенная онтологическая модель корреляции данных безопасности с условно-статичным содержимым ориентирована на использование взаимосвязей как между разнотипной информацией: уязвимостями, эксплойтами, слабостями, шаблонами атак и др., так и между различными источниками однотипной информации, например, между записями об уязвимостях в различных базах. Основным аспектом установления факта наличия связи между разнотипными данными безопасности является реализация той или иной сущности (эксплойт реализует уязвимость, уязвимость реализует слабость и т.д.). В свою очередь связь между экземплярами однотипной информации безопасности из различных источников строится за счет пересекающихся ссылок на ресурсы их описания.

2. Модель корреляции данных с динамичным содержимым (событий безопасности) опирается на извлечение характеристик объектов инфраструктуры из свойств событий. Множество различных комбинаций характеристик и их типов, а также связей между свойствами по различным показателям эквивалентности формируют конкретную модель неопределенной ЦИ КФС.

3. Разработанная комплексная методика корреляции информации и событий безопасности отличается адаптацией к неопределенной ЦИ КФС за счет выявления различных типов объектов и их иерархии на основе структурного анализа журналов событий. Дальнейшая обработка данных опирается на исследование поведения экземпляров объектов конкретных типов, а также связи информации с условно-статичным и динамичным содержимым.

Обоснованность и достоверность представленных научных положений обеспечивается тщательным анализом состояния исследований в предметной области, подтверждается согласованностью результатов, полученных при экспериментах, успешной апробацией на ряде научных конференций всероссийского и международного уровня, и публикацией в ведущих рецензируемых научных изданиях.

Теоретическая и практическая значимость результатов исследования. Разработанные модели и комплексная методика корреляции данных безопасности КФС развивают теоретические положения в данной области и позволят повысить эффективность предварительной оценки вероятности наступления инцидента за счет адаптивного определения ЦИ и идентификации типов активов, а также анализа поведения отдельных объектов. Предложенная методика должна стать основой подсистемы корреляции в SIEM-системах следующего поколения. Применимость результатов исследований определяется необходимостью проактивного мониторинга подобными системами состояния безопасности КФС для своевременного предотвращения потенциальных инцидентов и (или) минимизации ущерба в случае их неизбежности. С учетом критичности различных КФС, а также их разнообразия, предлагаемый подход корреляции данных имеет обширную область применения.

Реализация результатов работы. Отраженные в научно-квалификационной работе исследования проведены в рамках следующих научно-исследовательских работ: гранта РФФИ № 16-37-00338-мол_а, гранта РНФ № 15-11-30029, гранта Президента Российской Федерации № МК-314.2017.9, проектов Минобрнауки России № 14.604.21.0137, № 14.604.21.0033, № 14.616.21.0028 и № 14.604.21.0147, проекта 2017-2018 гг. НИР-ФУНД Университета ИТМО № 717075 «Методы, модели, методики, алгоритмы, протоколы и приложения для обеспечения информационной безопасности киберфизических систем» и др.

Апробация результатов работы. Основные положения и результаты работы докладывались на научных конференциях: 26-я международная конференция по параллельной, распределенной и сетевой обработке PDP-2018 (Кембридж, Великобритания, 2018); 9-я международная конференция по интеллектуальному сбору данных и передовым вычислительным системам IDAACS-2017 (Бухарест, Румыния, 2017); 23-я международная конференция по параллельной, распределенной и сетевой обработке PDP-2015 (Турку, Финляндия, 2015); 11-й международный симпозиум по интеллектуальным распределенным вычислениям IDC-2017 (Белград, Сербия, 2017); XX международная конференция по мягким вычислениям и измерениям SCM-2017 (Санкт-Петербург, 2017); 23-я и 24-я общероссийская научно-техническая конференция «Методы и технические средства обеспечения безопасности информации» (Санкт-Петербург, 2014-2015гг); IX и X Санкт-Петербургская межрегиональная конференция «Информационная безопасность регионов России (ИБРР-2015,2017)» (Санкт-Петербург, 2015,2017гг.); часть 7-й и 9-й Российской мультиконференции по проблемам управления – конференция «Информационные технологии в управлении (ИТУ-2014,2016)» (Санкт-Петербург, 2014,2016гг.); XIV и XV Санкт-Петербургская международная конференция «Региональная информатика (РИ-2014,2016)» (Санкт-Петербург, 2014, 2016гг.); VI международная научно-техническая и научно-методическая конференция «Актуальные проблемы инфотелекоммуникаций в науке и образовании (АПИНО-2017)» (Санкт-Петербург, 2017); всероссийский форум «Система распределенных ситуационных

центров как основа цифровой трансформации государственного управления (СРСЦ-2017)» (Санкт-Петербург, 2017), 4-я и 6-я международная конференция по практической безопасности «Positive Hack Days (PHD-4,6)» (Москва, 2015, 2017гг.).

Публикации. По материалам диссертационной работы опубликовано более 30 работ, в том числе 8 [3-10] – в рецензируемых изданиях из перечня ВАК («Информационно-управляющие системы», «Безопасность информационных технологий», «Проблемы информационной безопасности. Компьютерные системы», «Известия высших учебных заведений. Приборостроение», «Труды СПИИРАН», «Вопросы кибербезопасности»), 8 – в изданиях, индексируемых в международных базах Scopus и Web Of Science, и 7 свидетельств о государственной регистрации программ для ЭВМ.

Структура и объем научно-квалификационной работы. Данная работа включает введение, три главы, заключение, список литературы (74 наименования). Объем работы – 107 страниц машинописного текста; включает 21 рисунок и 7 таблиц.

Кратное содержание работы. В первой главе детально описываются сферы применения КФС как сложных систем управления, а также подробно рассматриваются проблемы обеспечения безопасности систем данного класса. Рассматривается место и роль процесса корреляции в SIEM-системах. Приводится общая схема следования потоков исходных данных для процесса корреляции, классы и виды информации безопасности. Указываются основные отличия между данными с условно-статичным и динамичным содержимым, а также причины подобного разделения. Подробно рассматриваются этапы процесса корреляции, а также отдельные методы корреляции и их общая классификация. Выдвигаются требования к комплексной методике корреляции данных безопасности по аспектам оперативности, масштабируемости и ресурсопотребления, а также основных задач, возлагаемых на процесс корреляции информации безопасности. Формулируются требования к исходным данным, отображающим ограничения

предлагаемой методики корреляции. В завершении главы приводится формальная постановка задачи научно-квалификационной работы.

Во второй главе рассматриваются теоретические положения, описывающие модель неопределенной ЦИ, на базе которой производится разработка моделей корреляции данных с условно-статичным и динамичным содержимым. Основу анализируемой ЦИ составляют множества информационных объектов, а также их типов и взаимодействий между ними. Определение множества типов событий направлено на преодоление структурной и конфигурационной неопределенностей ЦИ КФС, тогда как определение множества взаимодействий между типами информационных объектов ликвидирует функциональную неопределенность.

Описанная модель корреляции данных с условно-статичным содержимым разработана с помощью онтологического подхода, поскольку автоматизированное выявление семантических связей между подобной полуструктурированной информацией практически не выполнима. В свою очередь, модель корреляции данных с динамичным содержимым включает множество журналов безопасности, их типов событий и типов свойств событий. Такое представление данных позволило не только определять элементы модели ЦИ в рамках этапа адаптации, но и связать два класса информации (условно-статичной и динамичной) за счет комплексной методики корреляции событий и информации безопасности. В основе методики лежит структурный анализ журналов событий безопасности, позволяющий идентифицировать типы активов ЦИ КФС, их взаимосвязи, а также выделять отдельные активы для проведения поведенческого анализа каждого из них. Условно-статичная информация в данном случае выполняет роль частных характеристик информационных объектов ЦИ КФС.

В третьей главе подробно описывается опытный стенд и наборы данных, за счет которых производятся эксперименты по соответствию полученных результатов исследований предъявляемым требованиям. Рассматриваются показательные случаи результатов адаптации комплексной методики корреляции данных безопасности к неопределенной ЦИ КФС, а также их причины и варианты

преодоления коллизий различного рода. Оценивается возможность применения предлагаемого подхода корреляции к ЦИ условно-неограниченного размера. Полученные результаты подтверждают выполнение поставленной научной задачи, их теоретическую и практическую значимость, а также наличие инновационной составляющей подхода по отношению к существующим решениям.

Глава 1 Современное состояние проблемы корреляции данных для задачи обеспечения безопасности киберфизических систем

1.1 Задача обеспечения безопасности КФС

Киберфизические системы находят все большее применение в жизни человека как в сугубо промышленных направлениях, так и в личном пользовании. Наряду с такими сложными объектами, как: атомные станции, гидроэлектростанции, космические аппараты, самолеты и высокоскоростные железные дороги, КФС также применяются в данный момент при реализации концепций «Умный дом», «Умный город», «Умная дорога» и др.

Под КФС понимаются объекты синтеза кибернетических (информационных, вычислительных) технологий и реальных физических (технологических) процессов. Иными словами, данные системы являются результатом интеграции кибернетики и промышленных технологий, обусловленной главным образом автоматизацией процессов мониторинга и управления технологическими процессами с целью рационализации ресурсопотребления. Однако, информатизация технологических процессов, направленная на уменьшение роли участия в них человека, накладывает дополнительные требования к подобным системам. Таким образом, работа КФС должна сопровождаться обеспечением должного уровня точности, отказоустойчивости и безопасности. Возможность безусловного управления системами данного класса на любом из этапов процесса выполнения также является необходимым условием, которое должно быть реализовано еще на этапе проектирования.

В дальнейшем будут рассмотрены аспекты, характеризующие КФС, на основе которых *киберфизические инфраструктуры* (КФИ) будут рассматриваться как сложные динамические объекты. Также вопросы проактивного мониторинга состояния КФИ будут рассматриваться с точки зрения аспекта их безопасности.

Основными средствами обеспечения автоматизированного выполнения технологических процессов КФС являются автоматизированные системы управления технологическими процессами (АСУ ТП). В связи с этим, актуальность задачи мониторинга безопасности КФИ непосредственно зависит от уровня защищенности данных средств. В ходе исследований по анализу рынка АСУ ТП экспертами компании Positive Technologies были рассмотрены уязвимости компонентов порядка 500 производителей автоматизированных систем. В результате исследований с 2012 года было выявлено 743 уязвимости в АСУ ТП (рисунок 1) [11].

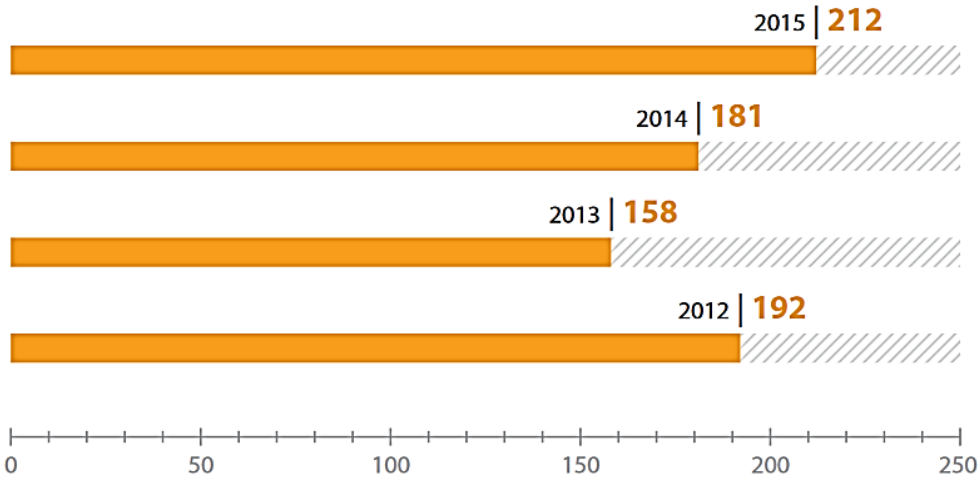


Рисунок 1 - Общее количество уязвимостей,
обнаруженных в АСУ ТП с 2012 года [11]

Лидерами в рейтинге наиболее уязвимых компонентов АСУ ТП в 2015 году (как и в 2012 году) являются продукты Siemens, Schneider Electric и Advantech. В категорию «Другие» вошел 81 производитель: их продукты имеют меньше 5 уязвимостей (рисунок 2) [11].

Наибольшее количество уязвимостей было выявлено в SCADA-компонентах и компонентах человеко-машинного интерфейса (ЧМИ), программируемых логических контроллерах (ПЛК), терминалах удаленного доступа и управления (ТУД), сетевых устройствах промышленного назначения и инженерном программном обеспечении (рисунок 3) [11].

Результаты, полученные в рамках исследований компании Positive Technologies, подтверждают актуальность задачи мониторинга безопасности в КФС.

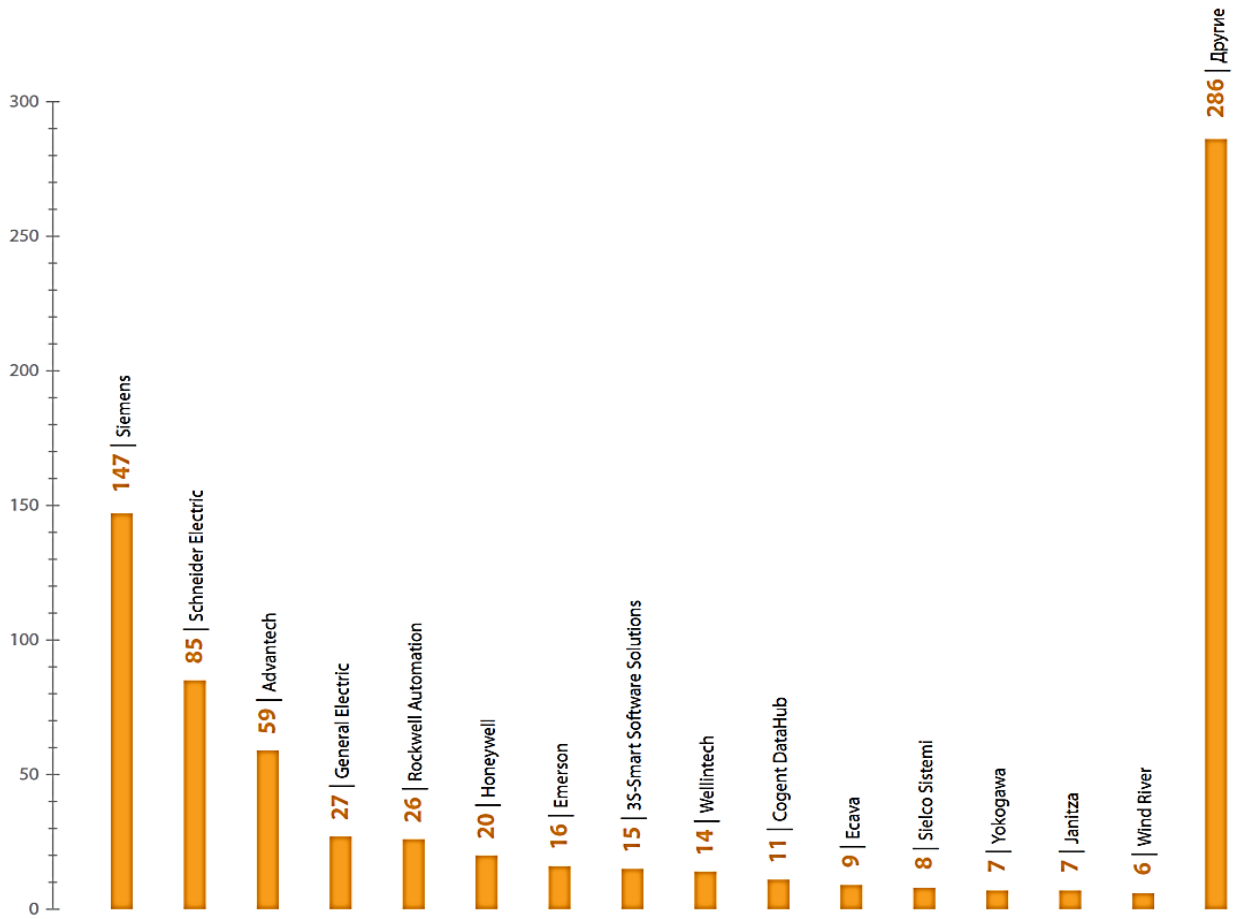


Рисунок 2 - Количество уязвимостей в компонентах АСУ ТП различных производителей [11]

Одним из наиболее показательных примеров вредоносного воздействия на КФИ является целевая киберфизическая атака Stuxnet, направленная на завод по обогащению урана с целью срыва ядерной программы Ирана, произошла в июне 2010 года [12]. Особенности данной атаки с точки зрения воздействия на технологический процесс заключаются в: получении слепка определенной промышленной системы управления, изменении кода в ПЛК Siemens для потенциального осуществления саботажа системы, скрытии измененного кода на ПЛК (по существу представляет собой rootkit для ПЛК) [13].

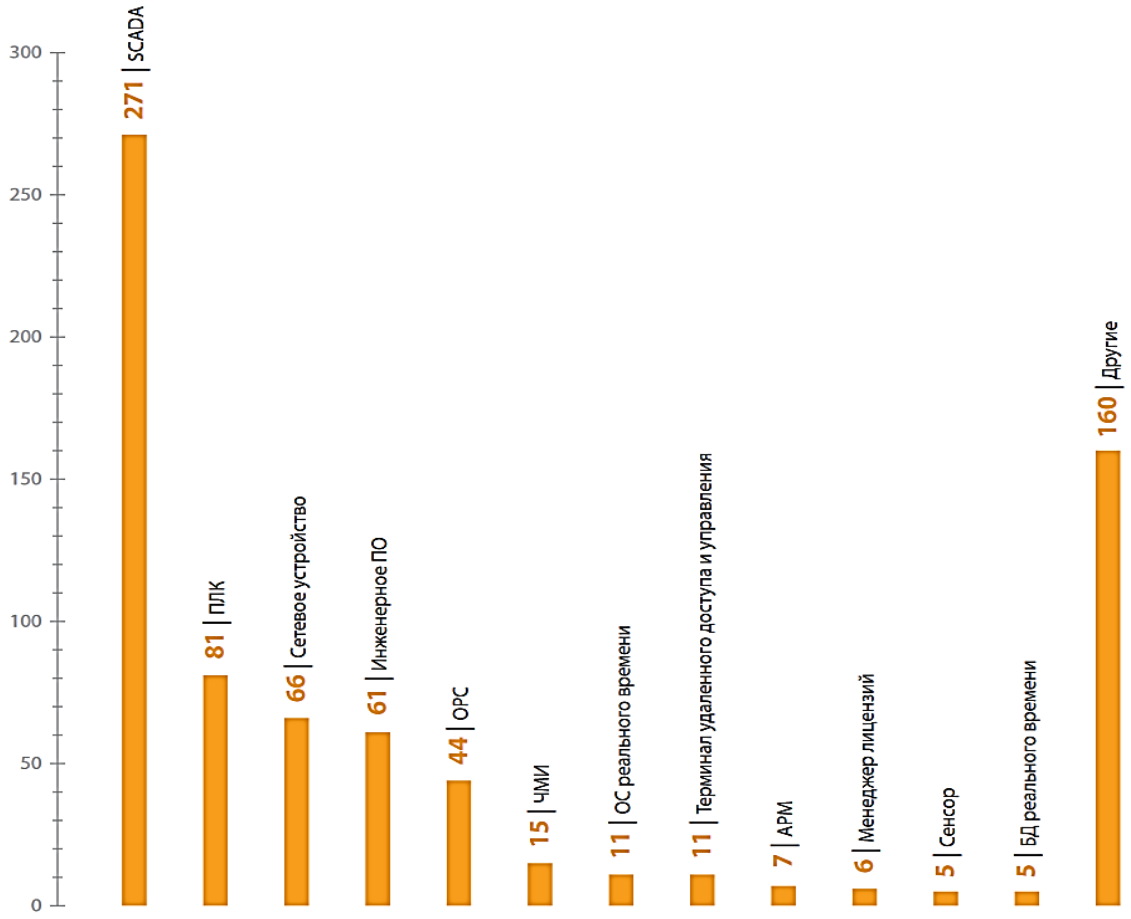


Рисунок 3 - Уязвимости в различных компонентах АСУ ТП [11]

Stuxnet стал первым компьютерным червем, который способен перехватывать и модифицировать поток данных между программируемыми логическими контроллерами марки SIMATIC S7 и рабочими станциями SCADA-системы SIMATIC WinCC фирмы Siemens. Вредоносная программа может использоваться злоумышленниками для несанкционированного сбора данных и диверсий в АСУ ТП промышленных предприятий, электростанций, аэропортов и пр. [12]. Именно после обнаружения атаки Stuxnet в мировом сообществе стали уделять внимание задаче мониторинга и обеспечения киберфизической безопасности КВО. Также данная атака была родоначальницей класса целевых атак (Advanced Persistent Threat, APT).

Сравнительно недавно стали известны подробности целевой атаки на энергетическую сеть Украины в декабре 2015 года. Данная атака была направлена на ряд электро-распределяющих центров Прикарпатья и началась задолго до того, как была осуществлена главная цель – отключение подстанций. Первым этапом

было осуществление фишинговой рассылки сотрудникам компаний, обслуживающих центры управления, после чего злоумышленники получили удаленный доступ к корпоративной сети 2-х из 3-х таких центров. Затем, атакующие, тщательно изучив особенности инфраструктур центров управления, произвели обход фаервола (межсетевого экрана), подмену прошивок SCADA-систем, отключение системы резервного питания для обесточивания самих центров и, наконец, отключение около 60 электроподстанций, что привело к потере электроэнергии 230 000 человек. Восстановление подачи электроэнергии было затруднено тем, что злоумышленники заблокировали доступ к системе управления с самой подстанции, а также тем, что сами диспетчерские пункты в центрах управления были некоторое время обесточены. Также данная атака сопровождалась TDoS-атакой на телефонную сеть call-центров, обслуживающих компанию поставщика электроэнергии, для невозможности сообщения жителями о потере электричества и соответственно увеличения времени реагирования на инцидент. Оперативное решение сложившейся ситуации стало возможным только из-за того, что в системах управления подстанциями была возможность ручного управления [14]. Стоит отметить, что такое же оборудование используется и на электроподстанциях США, только в некоторых случаях без возможности ручного управления, что только усугубило бы последствия подобной атаки. Также в СМИ неоднократно заявлялось об атаках на электроподстанции США в 2003 и 2008 году, в результате которых обесточенными в течение суток были порядка 50 000 человек, однако, официальные представители компаний-поставщиков электроэнергии и лица, занимающиеся расследованием данных инцидентов, всячески отрицали причастность к инцидентам киберфизического вредоносного воздействия.

Также стоит обратить внимание, что в настоящее время доказана возможность и опубликованы подробности выполнения вредоносного кода на ПЛК [15], а также существует несколько симуляторов (виртуальных лабораторий) по взлому КФИ КВО со специально написанным фрэймворком для достижения поставленных целей, например, атаки на химический завод [16].

Исходя из общепризнанной терминологии, *система* – это целостное образование, обладающее свойствами, не сводящимися к свойствам входящих в это образование взаимосвязанных (взаимодействующих) элементов (компонентов, частей, объектов, подсистем и т.п.). В свою очередь, *сложная система* – это система, познание (изучение) которой требует совместного привлечения разнотипных моделей, многих теорий, а в некоторых случаях, многих научных дисциплин (организации междисциплинарных исследований)[17]. Исходя из приведенного определения очевидно, что КФС больше относятся к сложным системам, чем к простым, поскольку помимо информационных технологий для обеспечения вычислительных процессов в них применяются технологии из других прикладных дисциплин.

Другой, более ранний термин сложных систем сформулировал Л.А. Растригин, выделивший основные черты (но не формальные признаки) сложной системы [18]:

1) Отсутствие математического описания сложной системы и необходимость в нем для выполнения задачи управления. В действительности, получение общей подробной математической модели КФС является важнейшей задачей, решаемой на стадии проектирования. Однако, для разработки универсальных моделей, рассчитанных как для статических, так и для изменяющихся во времени КФС, требуется особый подход. Также возможно, что создание данных моделей практически невозможно ввиду сильной разнородности сфер применения КФС. Именно поэтому для решения задач мониторинга состояния сложных систем и объектов данного класса на начальном этапе применяется принцип декомпозиции, когда разрабатываются модели для описания отдельных функциональных процессов системы, а затем, композиция полученных моделей в общую модель. Таким образом, задача мониторинга безопасности КФИ является обособленной, однако результаты ее выполнения будут находить свое применение в комплексной задаче обеспечения работоспособности всей системы.

2) Стохастичность поведения сложных систем. Данная характеристика отражает степень непредсказуемости последующих макро-состояний даже при наличии достаточной информации о текущем состоянии. В данном случае, КФС в целом имеют более стохастичное поведение, чем вычислительные или прикладные системы (элементы системы) по отдельности. Данный факт обусловлен как большим множеством состояний переходов, так и расширенным множеством воздействий внешней среды. Применение КФС главным образом нацелено на минимизацию роли участия в них человека для повышения эффективности работы всей системы, что несомненно влечет к снижению фактора человеческой ошибки при управлении. Но, не смотря на это, возрастает вероятность агрессивных воздействий на КФС ввиду объединения виртуальной вычислительной (информационной) внешней среды с физической внешней средой. Например, наряду с возможностью возникновения природных катаклизмов, негативно отражающихся в основном на выполнении технологических процессов, следует также учитывать возможные кибератаки и сбои, влекущие за собой нарушение работы вычислительных процессов и всей системы в целом.

3) «Нетерпимость» сложной системы к управлению, отражающаяся в изменении ее самостоятельного поведения. Данная черта характерна для КФС, в которых технологические процессы стремятся к саморегулированию в равновесное состояние. Основой данных процессов как правило являются физические процессы с динамически изменяемыми состояниями внешних сред.

4) Нестационарность сложной системы, отражающая эволюционные изменения системы с течением времени. Тривиальным примером нестационарного поведения сложных систем является амортизация (старение) элементов системы. В КФС данное свойство имеет место быть при использовании высокотехнологичного и высокоточного оборудование для выполнения технологических процессов.

5) Невоспроизводимость экспериментов над сложными системами. Данная черта присуща КФС, нарушение корректной работы которых может привести к

необратимым катастрофическим последствиям. Для преодоления данного ограничения разрабатываются теоретические модели состояний сложных объектов, которые включают предаварийные и аварийные состояния, однако проверить работу системы при переходе в данные состояния реально не представляется возможным ввиду возникновения большого риска причинения ущерба системе и окружающей среде. Один из возможных способов оценки качества таких моделей заключается в проведении имитационного моделирования, которое направлено на определение поведения системы при нежелательных и недопустимых состояниях.

Таким образом, было установлено, что КФС в общем представлении являются сложными. Для задачи обеспечения мониторинга безопасности КФС используется анализ их инфраструктур (КФИ), которые включают: множество датчиков (сенсоров), сообщающих о текущем состоянии и (или) его изменении для каждого элемента контроля. В данном случае, необходимой информацией являются не только показания источников, размещенных в среде технологических процессов, но и события, происходящие в среде вычислительных (информационных) процессов.

В свою очередь С. Бир привел свою классификацию систем, в которой выделяются детерминированные и вероятностные системы, а также простые, сложные и очень сложные системы [19]. Стоит отметить, что детерминированная система в рамках данной классификации не может быть очень сложной.

При замкнутой архитектуре, когда внешние кибернетические (информационные) факторы воздействия отсутствуют, а физические факторы сведены к минимуму, КФС может быть отнесена к сложным детерминированным системам. Однако, для задачи мониторинга состояния безопасности КФС, когда основными рассматриваемыми и допускаемыми факторами воздействия на защищенность являются внешние кибернетические и физические факторы, данные системы будут относиться к сложным и очень сложным вероятностным системам.

Также, одной из важнейших характерных черт сложных систем является наличие у системы таких возможных состояний, которые не присущи отдельным входящим в нее элементам. В рамках задачи мониторинга безопасности КФС, глобальным (общим, комплексным) состоянием системы может выступать текущий уровень защищенности, на который непосредственно влияет уровень защищенности отдельных компонентов системы. Однако, состояние отдельно взятого элемента не является равнозначным общему состоянию системы. Стоит отметить, что на определение глобального состояния КФС влияют состояния элементов технологических процессов, не имеющих функции определения уровня защищенности, что усложняет задачу определения общего состояния системы.

1.2 Место роль процесса корреляции для мониторинга и управления безопасностью. Исходные данные процесса корреляции

Место и роль процесса корреляции. Процесс корреляции данных в SIEM-системах занимает одну из основополагающих ролей. Наряду с процессами оценки рисков, моделирования угроз и расследования инцидентов, наличие которых в конкретном решении зависит от реализации, система корреляции является безусловно необходимым для выполнения задач мониторинга и управления безопасностью.

Для постановки задачи исследований, первоначально, следует определить место и роль процесса корреляции в SIEM-системах. Считается, что процесс корреляции направлен на (1) определение взаимосвязей между разнородной информацией безопасности; (2) определение информационных объектов целевой инфраструктуры, их характеристик и иерархии; (3) группировка низкоуровневых событий в высокоуровневые мета-события; и (4) выявление потенциальных инцидентов и предупреждений безопасности на основе анализа поведения разных объектов инфраструктуры [20]. Таким образом, процесс корреляции производит обработку данных от поступления их из гетерогенных источников и до формирования отчета о текущем состоянии защищенности анализируемой

инфраструктуры. Стоит также отметить, что процесс корреляции является непрерывным и должен быть рассчитан на выполнение в реальном масштабе времени.

Как правило в процессе корреляции данных выделяются следующие этапы (рисунок 4)[7,21]:

- Нормализация (приведение собираемых данных к единому формату).
- Анонимизация (преобразование данных, ограниченных от нежелательного разглашения).
- Фильтрация (отсеивание малозначительных или бесполезных данных).
- Агрегация (получение новых данных с использованием различных функций агрегирования, таких как average(), count(), min(), max() и др.).
- Анализ (нахождение закономерностей появления и зависимостей (в том числе скрытых) между событиями безопасности).
- Корреляция (определение взаимосвязей между экземплярами событий и их групп).
- Ранжирование (оценка результатов корреляции по определенным признакам).
- Приоритезация (вычисление степени важности результатов процесса корреляции).

Более детально каждый из этапов будет рассмотрен в разделе 1.3.

Исходные данные для процесса корреляции. В качестве входных данных процесса корреляции могут выступать различные источники информации (внутренние и внешние): сенсоры (датчики) измерений, агенты сбора данных, журналы событий, конфигурации объектов инфраструктуры и многие другие. Общая схема использования различных источников данных приведена на рисунке 5, в котором базы программно-аппаратного обеспечения (ПАО) хранилища содержат информацию об идентификационных характеристиках установленных программно-аппаратных средств в ЦИ.

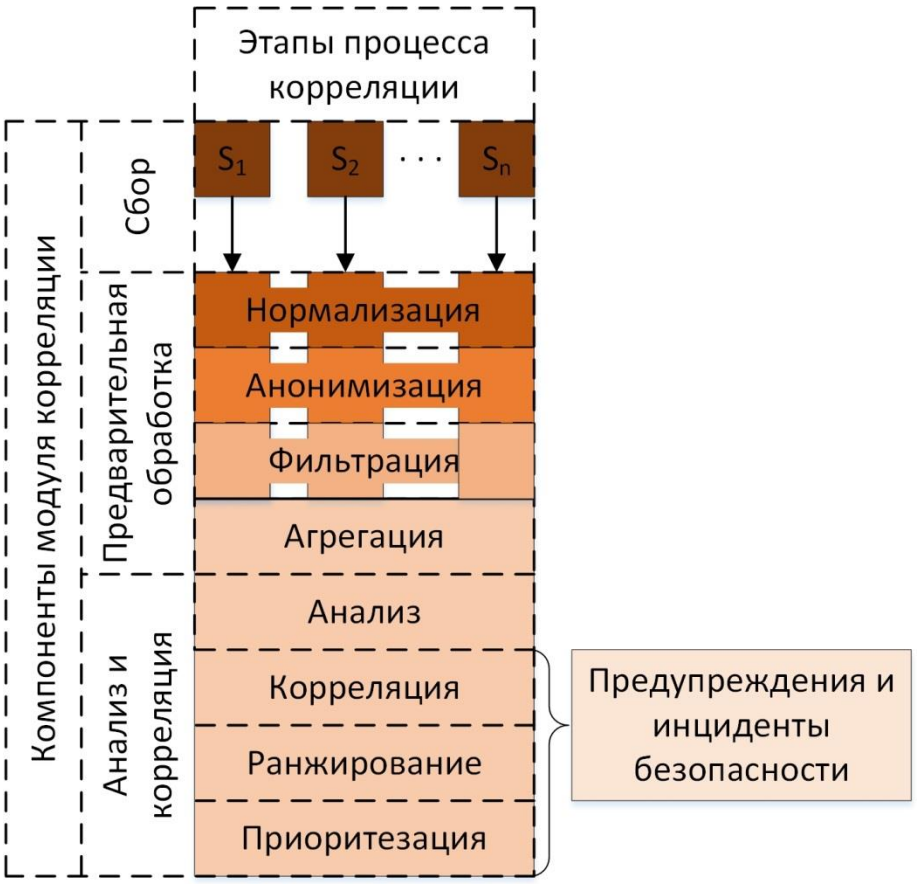


Рисунок 4 – Общая схема этапов процесса корреляции

На данной схеме входные (сырые) данные представлены внутренней информацией с динамичным содержимым и внутренней и внешней информацией с условно-статичным содержимым. Данное разделение необходимо ввиду сложности корреляции в одном процессе информации из разных категорий, главным различием которых является привязка к масштабу времени (для динамического содержимого). Также представленная схема включает средства защиты, осуществляющие промежуточную обработку входной информации и генерирующие более высокоуровневые события. Однако, связь источников данных с приведенными средствами не фиксирована, то есть использование того или иного источника конкретным средством зависит от его реализации. Таким образом, исходными данными для выполнения процесса корреляции являются разнородные и разноуровневые события и данные безопасности с условно-статичным содержимым.

Рисунок 5 - Общая схема использования входных данных для процесса корреляции в SIEM-системах

Ранее анализ источников информации безопасности и их форматов широко рассматривался в литературе [22-24]. В работе [3] рассматриваются источники баз уязвимостей и их форматов представления. В работе [25] описывается формат SCAP (Security Content Automation Protocol), включающий несколько видов информации безопасности: уязвимости, конфигурации, платформы, метрики уязвимостей и другие. В [26] подробно рассматривается база CAPEC (Common Attack Pattern Enumeration and Classification) [27] и примеры её использования для задачи оценки защищенности компьютерных инфраструктур.

Основное отличие между исходной информацией с условно-статичным и динамичным содержимым заключается в характере ее изменчивости. Данные с условно-статичным содержимым после добавления в какое-либо хранилище как правило не изменяются (за исключением редких случаев корректировки и дополнения информации). Данные с динамичным содержимым, даже будучи жестко-структурированными, отображают характеристики описываемых аспектов безопасности в реальном масштабе времени, то есть их непосредственное

использование через значительный интервал времени будет малоэффективно. В текущем рассмотрении информации безопасности к данным с динамичным содержимым относятся события, как результат действия или попытки к совершению действия, формируемый либо источником действия, либо системой его обработки, а также обладающим специфичными свойствами, описывающими само действие. Также стоит внести ясность в понятия вид и источник информации безопасности. Под видом понимается отдельный класс объектов, характеризующих определенный аспект безопасности (уязвимость, слабость и т.д.). Под источником подразумевается конкретная база, предоставляющая сведения об отдельном виде информации.

Наиболее обширным видом информации безопасности, как по общему объему, так и по количеству источников, являются описания *уязвимостей*. На данный момент наиболее распространенной базой уязвимостей является CVE (Common Vulnerabilities and Exposures) [27]. Ключевыми полями описания уязвимостей в формате CVE являются: уникальный идентификатор источника; текстовое описание уязвимости; ссылка на другие источники описания уязвимости; дата публикации и статус уязвимости (кандидат или проверенная уязвимость). Расширенной версией CVE является база NVD (National Vulnerability Database) [28], содержащая дополнительные поля: список и конфигурацию программно-аппаратного обеспечения (в формате CPE [27]), подверженного уязвимости; оценку уязвимости в системе CVSS (Common Vulnerability Scoring System) [29]; результат эксплуатации уязвимости и идентификатор слабости (CWE) [27], используемой для реализации уязвимости. Последнее поле не является обязательным и зачастую имеет недостаточно точную характеристику, так как преимущественно используются только верхнеуровневые классификаторы слабостей. Прочие источники описания уязвимостей, такие как OSVDB [30], X-Forse [31] и BugTrack [32], содержат схожие по смыслу поля. Главным недостатком данных баз является неунифицированное представление уязвимых продуктов и отсутствует описание их конфигурации. К достоинствам использования различных источников описания уязвимостей следует отнести

расширение доступной информации, что может быть полезным для формирования гибридного хранилища данных безопасности.

Базы (словари) программно-аппаратных средств используются для выявления существующих в системе уязвимостей и также имеют определенный формат описания записей. Так, упомянутый выше словарь CPE (Common Platform Enumeration) [27] версии 2.2 имеет следующие поля записи: тип; имя производителя; название продукта; версия; модификация; редакция; язык. Версия 2.3 дополняется полями: архитектура; аппаратная платформа; программная платформа. Данное расширение должно более точно определять различные программно-аппаратные средства в словаре, однако, ввиду особенностей формата записи в словаре, нередко возникают проблемы с однозначной идентификацией продукта. Более совершенный вид словаря программно-аппаратных средств имеет формат CVRF (Common Vulnerability Reporting Framework) [33] с иерархичным представлением различных разновидностей продуктов и добавленными полями: номер сборки; спецификация и другими. Однако в указанном формате словарь продуктов на настоящий момент опубликован не был.

Информация об *эксплойтах* представляет собой описание практической реализации конкретной уязвимости, и как правило содержит поля: уникальный идентификатор источника; программно-аппаратные средства, в которых реализуется уязвимость; детали эксплуатации; программный код, направленный на нарушение безопасности.

Базы конфигураций содержат информацию о корректной и безопасной настройке конкретных программных средств. На данный момент в формализованном виде существует единственная открытая база подобного вида информации безопасности – «Общее перечисление конфигураций» (Common Configuration Enumeration, CCE) [27]. Формат указанной базы содержит поля: уникальный идентификатор; текстовое описание настраиваемого компонента платформы; неформатированное имя платформы; параметры настройки конфигурации; технические детали механизма настройки конфигурации, указывающие на расположение опциональных полей (путь в файловой системе,

имя параметра в конфигурационном файле, ветка реестра и др.); ссылки на источники рекомендаций по настройке текущей конфигурации (как правило, на сайте разработчика платформы).

Шаблоны атак являются ключевым видом информации безопасности для обнаружения и предотвращения атакующих действий на распределенные и связанные сетью объекты защищаемой инфраструктуры. В качестве основного источника описания указанного вида информации выступает база CAPEC [27] со следующими полями формата представления: уникальный идентификатор источника; идентификаторы примеров уязвимостей; идентификаторы слабостей, используемых при атаке; текстовое описание атаки; ссылки на дополнительное описание отдельных этапов или деталей атаки (используемые уязвимости, эксплойты, программно-аппаратное обеспечение и др.).

Информация о *слабостях* ЦИ с точки зрения безопасности на данный момент представляется в виде классификации, содержащей в иерархическом виде следующие поля: категории; классы; базовые слабости и варианты. Каждый элемент имеет уникальный идентификатор общепринятого источника CWE (Common Weaknesses Enumeration) [27].

Хранилищами для сбора *событий* являются журналы операционных систем, приложений, сервисов и других всевозможных источников и сенсоров. В связи с этим, форматы описания событий в разных журналах значительно отличаются друг от друга. Между тем, практически все форматы представления событий обладают и общими полями (свойствами): дата и время создания события; глобальный тип события (информационное, предупредительное, завершенное с ошибкой, аудит успеха, аудит отказа); собственный тип, указывающий на конкретное действие, которое описывается в событии; источник события (сенсор, приложение и прочие) и некоторые другие. В отличие от информации с условно-статичным содержимым, отсутствующие значения некоторых полей в записях событий могут быть частично восстановлены. Например, собственный тип события в неявном виде может быть отнесет к определенной категории событий, не указанной в записи журнала. Подобная ситуация и с категориями свойств

событий. Задача хранения, обработки и дальнейшего использования событий как информации безопасности реализуется в подсистеме корреляции системы аналитической обработки информации и событий безопасности. При выполнении этапа нормализации производится соотнесение форматов журналов событий между собой, а при непосредственной корреляции – формирование высокоуровневых событий в гибридном хранилище информации безопасности.

1.3 Этапы процесса корреляции. Методы корреляции данных безопасности и их классификация

Работа компонента корреляции, в широком смысле, направлена на обнаружение атак, вредоносной активности и нарушений политики безопасности. В узком смысле, компонент корреляции предназначен для поиска связей, зависимостей и причинно-следственных отношений между событиями безопасности и другой информацией безопасности. Данный компонент выполняет как функции корреляции событий, так и их пред- (пост-) обработку в зависимости от конкретной реализации системы.

Учитывая необходимость оперирования событиями и информацией безопасности, компонент корреляции можно рассматривать с разных точек зрения. Корреляция представляется либо с точки зрения рассмотрения ее как процесса, либо с точки зрения множеств входных и выходных данных, преобразуемых внутри системы корреляции. В первом случае, корреляция представляет собой последовательность операций над событиями, специально определенную для получения конкретного решения и являющуюся непрерывной относительно работы всей системы. Во втором случае, корреляция задается через множество типов событий [34], преобразуемых таким образом, что несколько событий могут образовывать одно более сложное событие и восприниматься системой как неделимое [35-37].

Стоит отметить, что в процесс корреляции можно представить с использованием двух основных типов операций (функций): (1) комбинирования

данных безопасности; (2) идентификации и удаления (или обозначения) ложных или бесполезных данных безопасности [21]. Важными операциями корреляции также являются представление событий безопасности и обучение (переобучение) системы корреляции во время работы. Вместе с тем, данные типы операций на разных этапах процесса корреляции выполняют отличающиеся по своему характеру действия.

Наряду с общей схемой процесса корреляции, большую значимость при обработке событий безопасности имеют используемые методы поиска взаимосвязей и их параметров над множеством входных данных (событий и информации безопасности). За последнее время, в рамках исследования процесса и методов корреляции данных безопасности предложены различные методики обработки разнородных данных и преобразования в процессе корреляции низкоуровневых событий к высокоуровневым, а также рассмотрены возможные схемы, описывающие сам процесс корреляции [38-49].

Обзор релевантных работ. В [38] предлагается разбиение процесса корреляции на выполняемые задачи. Выделяются следующие составляющие процесса: *сжатие* (*compression*), *счёт* (*count*), *подавление* (*suppression*), *логическая замена* (*boolean*) и *обобщение* (*generalization*). Под *сжатием* подразумевается преобразование в одно событие безопасности нескольких одинаковых событий. *Счёт* представляет собой замену похожих событий безопасности одним новым событием, а *подавление* – осуществление задержки обработки событий безопасности с низким приоритетом до окончания обработки события безопасности с более высоким приоритетом. Процесс *логической замены* заключается в преобразовании некоторого множества событий в новое событие, удовлетворяющее определённому логическому шаблону. В результате выполнения *обобщения* производится перевод события безопасности к высокоуровневому представлению (суперклассу) для удовлетворения необходимой важности уведомления. Несомненным достоинством работы является возможность добавления статистических методов в описанную модель корреляции событий безопасности, несмотря на то, что она основана на строго

детерминированных подходах. Недостатки представленной работы заключаются в отсутствии среди указанных задач элемента *предупреждения ошибок (fault prediction)* и элемента *предупреждения нарушений (preventive maintenance).*

В [39] описываются методы корреляции предупреждений для управления сетевыми сбоями. Весь процесс корреляции делится на четыре этапа: (1) *фильтрация* (*filtering*), (2) *корреляция* (*correlation*), (3) *идентификация сбоя* (*fault identification*) и (4) *коррекция* (*correction*). Стоит отметить, что для каждого этапа в этой работе выделен наиболее применимый способ выполнения, а именно: *экспертная система* (expert system), *нейронная сеть* (neural network) на первых двух этапах и *вывод на основе прецедентов* (*case-base reasoning*) на третьем и четвертом. В предложенной схеме процесса корреляции используется библиотека прецедентов (case library) и выделяется цикл между корреляцией и идентификацией сбоев.

В [40] также приводится собственная классификация систем корреляции, согласно которой выделяются следующие типы систем*:* (1) *системы, основанные на правилах* (*rule-based*); (2) *системы на базе кодовых книг* (*codebook*) и (3) системы *с использованием интеллектуальных методов* (*artificial intelligence*). Однако, в рамках представленной работы, особенности и недостатки каждой из перечисленных типов систем не приводятся.

В [41] описываются две модели корреляции событий безопасности: (1) *причинно-следственная* и (2) *временная*. Данные модели основаны на определении соответствия между событиями и упорядочивании событий в хронологическом порядке соответственно. В работе считается, что в системах корреляции обычно производится сопоставление указанных моделей с топологией анализируемой сети. Вместе с этим, в статье приводится доказательство использования *причинно-следственной* модели корреляции в составе *временной* модели. Стоит отметить, что простота приведенных моделей процесса корреляции событий безопасности делает их неприменимыми в качестве механизма непосредственного описания взаимосвязей между событиями в реальных

системах, а сама работа нацелена скорее *не на пользователей*, а на *разработчиков* систем корреляции.

В [42] производится обзор работ в области корреляции предупреждений для систем обнаружения вторжений (Intrusion Detection System, IDS). В частности, рассматриваются этапы и операции процесса корреляции, описывается модель данных формата обмена сообщениями обнаружения вторжений (Intrusion Detection Message Exchange Format, IDMEF), а также приводится пример процесса корреляции для обнаружения типовой атаки. Процесс корреляции в данной работе условно делится на три этапа: (1) предобработка (*preprocessing*); (2) анализ предупреждений (*alarm analysis*); (3) корреляция предупреждений (*alarm correlation*). На втором этапе выделяются такие методы и этапы процесса корреляции, как измерение схожих признаков (*similarity measures*), кластеризация, интеллектуальный анализ данных (*data mining*), удаление, редукция и слияние. Стоит отметить, что в результате выполнения каждого из трех этапов формируются простые события, мета-события и сценарии атак соответственно, а по окончании выполнения процесса корреляции формируется отчет.

В [43] выделяется шесть этапов процесса корреляции: нормализация (*normalization*), агрегация (*aggregation*), корреляция (*correlation*), отсеивание ложных срабатываний (*false alert reduction*), анализ стратегии атаки (*attack strategy analysis*) и приоритезация (*prioritization*). Описываются четыре основных метода корреляции: (1) на основе сценариев атак; (2) ориентированного на правила; (3) статистического и (4) временного. Главное отличие [43] от аналогичных работ заключается в точном связывании этапов процесса корреляции с используемыми в них конкретными методами. В данной работе также выделены отдельные группы методов корреляции.

Работа [44] посвящена разработке сервис-ориентированного приложения корреляции событий и содержит описание основных применяемых методов. Особенностью данной работы является описание примеров реализации методов корреляции как на этапе проектирования, так и в ходе непосредственного функционирования разработанного прототипа. Выделяются следующие методы

корреляции: метод на основе моделирования (model-based reasoning, MBR); правило-ориентированный метод (rule-based reasoning, RBR); метод на основе кодовой книги (codebook); метод рассуждений на основе прецедентов (case-based reasoning, CBR); метод активного исследования (active probing). В работе перечисленные методы корреляции были сравнены по различным свойствам, определена актуальность использования каждого метода в автоматизированном сервисе корреляции. Например, свойство возможности сопровождения метода (*maintenance*) характеризует способность модификации метода. Свойство поддержки моделирования (*modeling*) отражает возможность моделирования метода. Под свойством надежности (*robustness*) понимается отказоустойчивость встроенных механизмов исправления ошибок. Свойство производительности (*performance*) отражает быстродействие метода на основе сложности используемых им алгоритмов. По выделенным свойствам наиболее выигрышным является метод рассуждений на основе прецедентов. Однако, в предлагаемой архитектуре сервиса корреляции, данный метод используется только при ошибках в результате работы ранее используемых методов (правило-ориентированном и на основе моделирования). Также отмечается, что ни один из методов не годится для исключительного (единственного) использования в разрабатываемом сервисе корреляции.

В диссертации [45] приведено описание модельно-методического аппарата для корреляции событий безопасности. Работа включает разбор этапов процесса корреляции, применяемых методов и типов операций, производимых над данными, различных систем корреляции событий безопасности и их сравнительный анализ, а также форматы описания событий. Важным элементом данной диссертации является оценка преимуществ и недостатков использования различных функциональностей и режимов работы системы корреляции, таких как самообучение и использование внешних знаний, реальное время и сохранение данных (с сохранением состояния и без), активный и пассивный режимы работы, централизованное и децентрализованное управление, глубокий анализ и поверхностное исследование. Несмотря на то, что данные пары

функциональностей и режимов работы обладают противоположными качествами, в определенных случаях целесообразно, чтобы система обладала обоими свойствами из одной пары. Данная работа также достаточно полно раскрывает детали процесса корреляции с точки зрения применяемых методов и подходов. Выделяются следующие методы корреляции: (1) *машина конечных состояний*; (2) *правила*; (3) *поведенческий анализ*; (4) *моделирование*; (5) *кодовая книга*; (6) *голосование*; (7) *явное обозначение ошибки*; (8) *граф зависимостей*; (9) *байесовская сеть*; (10) *нейронная сеть*. Данные методы по своей сущности имеют один или несколько формальных подходов, таких как теория графов, нечеткая логика, теория вероятности, математическая статистика, машинное обучение, интеллектуальный анализ данных и другие.

В [46] предлагается различать методы корреляции событий безопасности по следующим ортогональным критериям: (1) *способ корреляции событий безопасности*; (2) *уровень корреляции событий безопасности*; (3) *используемые форматы данных*. Авторы работы отмечают, что форматы данных получаемых пакетов, потоков и событий безопасности должны быть определяемыми. Приводится классификация способов непосредственной корреляции, которая различает *сигнатурные* и *бессигнатурные* (на основе обнаружения аномалий) алгоритмы. При этом, в категории бессигнатурных алгоритмов выделяются (1) подходы, основанные на *спецификации* (specification-based), выполняющиеся, как правило, в ручном или полуавтоматическом режиме, и (2) подходы, базирующиеся на *интеллектуальном анализе данных* (data-mining-based). В работе отмечается, что бессигнатурные алгоритмы могут выбираться в зависимости от анализируемого трафика. Если *нормальное поведение системы* используется в качестве *данных обучения*, то генерация событий безопасности будет соответствовать *несовпадению* наблюдаемых данных с шаблоном обучения. Если же *аномальная активность системы* используется в качестве *данных обучения*, то генерация событий безопасности будет соответствовать *совпадению* наблюдаемых данных с шаблоном обучения.

В [47-49] авторы выделяют пять подходов к процессу корреляции событий безопасности, основанных на: (1) *подобии* (сходстве) (*similarity based*); (2) *предопределении сценариев атак* (*predefined attack scenarios based*); (3) *многоуровневых вычислениях* (*multi-stage*) *на базе предпосылок и последствий*; (4) *использовании множества источников информации* (*multiple information sources*); (5) *фильтрации* (*filter based*).

Первый подход заключается в вычислении величины подобия двух событий безопасности на основе атрибутов, ассоциируемых с этими событиями. События, величина подобия которых достаточно велика, группируются.

Второй подход заключается в объединении в последовательность связанных этапов проведения атак на основе заранее определенных шаблонов сценариев атак. Данный подход применяется для получения агрегированного и более высокоуровневого взгляда на угрозы безопасности.

Третий подход основывается на формировании сценариев атак путём связывания отдельных этапов их проведения при условии, что один из этапов является необходимым условием для проведения другого.

Четвёртый подход направлен на приоритезирование и классификацию потоков событий безопасности в зависимости от источника информации о событии безопасности.

Пятый подход основан на удалении из процесса корреляции событий по заранее определённым правилам (фильтрам). Решение об удалении события из процесса корреляции принимается на основе значений одного или нескольких его атрибутов.

Похожая классификация приводится в [50]. Приведённая авторами модель корреляции событий безопасности состоит из двух частей: (1) подхода, *основанного на графах атак* (*an attack graph-based*) и (2) подхода, *основанного на подобии* (*similarity-based*). При этом первый подход используется для корреляции событий безопасности, вызванных известными атаками, а также для построения гипотез о вероятно необнаруженных или упущенных событиях безопасности. В

свою очередь, второй подход применяется для корреляции событий безопасности, вызванных неизвестными атаками, а также для уточнения известных графов атак.

В [21,46,51] выделяются и описываются отдельные этапы и уровни процесса корреляции.

В [21] авторы раскрывают методы и подходы к корреляции предупреждений в зависимости от фазы процесса, указывают на их достоинства и недостатки, а также спорные моменты. В работе процесс корреляции предупреждений представлен в виде этапов, которые преобразуют оповещения сенсоров в отчеты о вторжениях и направлены на разные аспекты процесса корреляции (рисунок 6).

Именно данная схема взята за основу во многих других работах (в том числе и в настоящей статье) для представления этапов процесса корреляции и их последовательности. В [21] рассматриваются методы корреляции предупреждений в системах обнаружения вторжений (Intrusion Detection System, IDS), которые также применимы в SIEM-системах, несмотря на то, что предупреждения являются только отдельным типом событий. Иными словами, за счет расширения типов обрабатываемых событий SIEM-системы расширяют системы обнаружения вторжений в рамках корреляции. В данной работе также производится структурное деление этапов процесса корреляции предупреждений в зависимости от их целей, таких как: сбор (*collection*), агрегация и верификация (*aggregation and verification*), анализ высокоуровневых структур (*high-level structures*), крупномасштабная корреляция (*large-scale correlation*), оценка (*evaluation*).

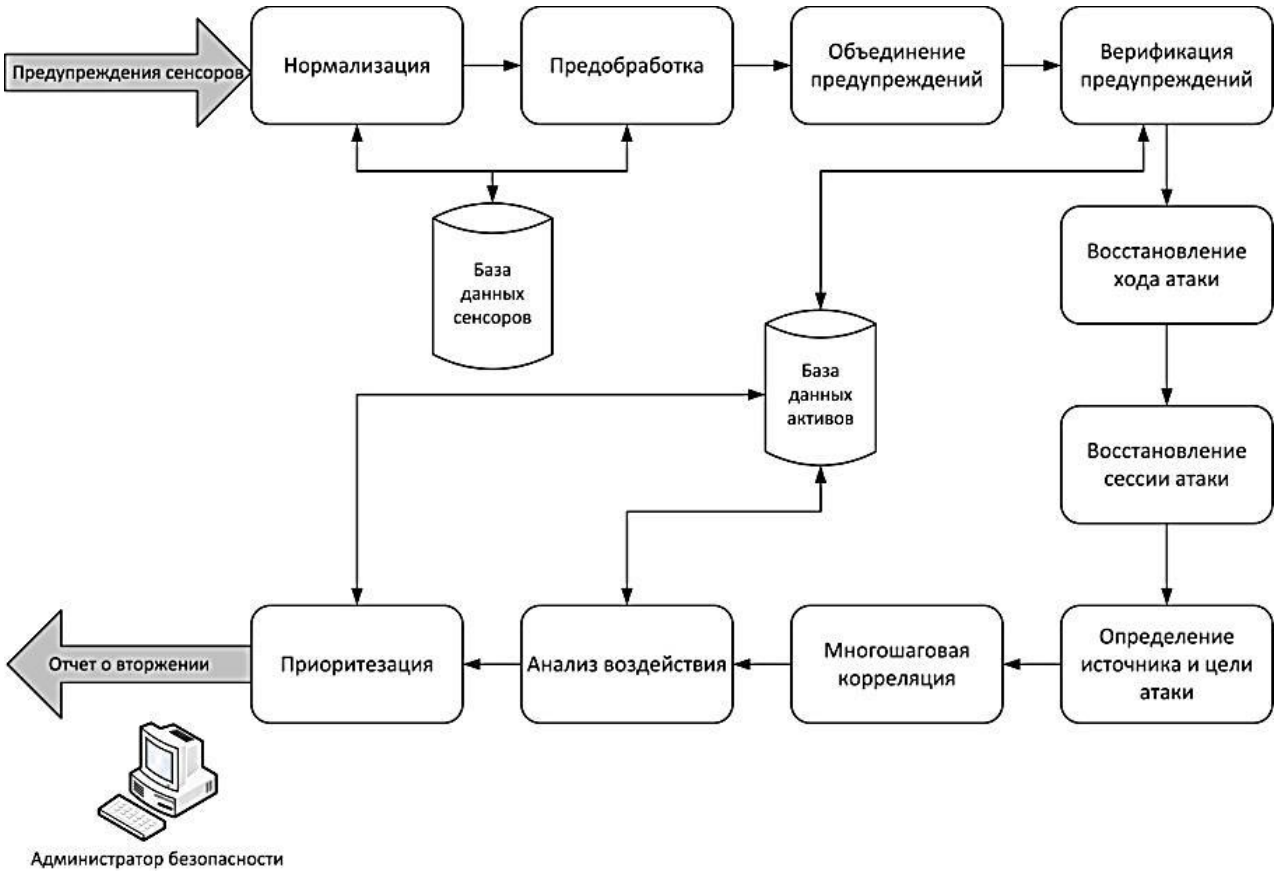


Рисунок 6 - Представление процесса корреляции предупреждений в [21]

В рамках детектирования атак приводятся следующие виды систем: (1) на основе злоупотреблений (*misuse-based*; в основу положена база знаний; атака идентифицируется при соответствии записи базы с параметрами входных данных); (2) на основе аномалий (*anomaly-based*; текущее состояние сравнивается с эталонным с помощью оценки вероятности отклонения).

В [51] процесс корреляции предупреждений представлен в виде логических блоков. Авторы выделяют следующие блоки: *нормализация данных* (*Data Normalization Unit*), *корреляция на основе фильтрации* (*Filter-based Correlation Unit*), *редукция данных* (*Data Reduction Unit*), *анализ намерений* (*Intention Recognition*) и *анализ воздействий* (*Impact Analysis*). Отличительной особенностью данной работы является представление модели процесса корреляции (рисунок 7), которая снижает количество обрабатываемых событий безопасности так рано, как это только возможно. Это осуществляется путём вывода из процесса корреляции незначимых или ложных событий безопасности ещё на начальных этапах процесса. Авторы ввели дополнительный компонент для работы с

некоррелируемыми данными. По итогам эксперимента на наборах данных DARPA 2000, авторам удалось добиться процента редукции 99,38 % (в среднем).

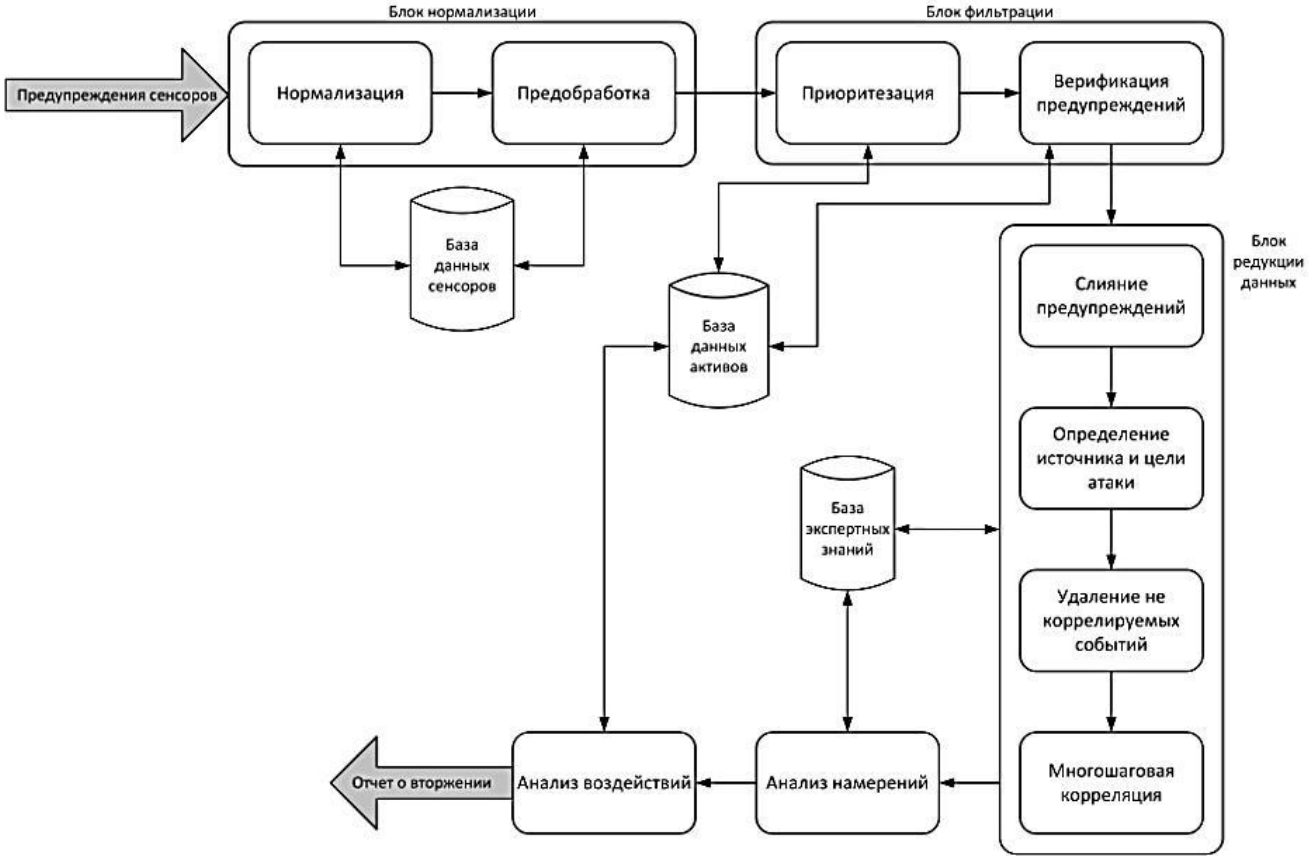


Рисунок 7 - Модель процесса корреляции событий безопасности [51]

Важно отметить, что данная модель не лишена недостатков. Во-первых, в блоке фильтрации, этапы верификации и приоритезации безусловно будут обрабатывать в том числе и дубликаты событий безопасности, так как модуль агрегации находится на более высоком уровне (блок редукции данных). Учитывая ресурсоёмкость процесса проверки событий безопасности на подлинность, данное решение подлежит дополнительному рассмотрению. Во-вторых, этап удаления из процесса корреляции событий безопасности данных, которые не могут быть коррелированы (*Удаление не коррелируемых событий*), не оставляет процессу права на ошибку. При увеличении среднего показателя редукции событий безопасности и облегчении дальнейшего анализа открытым остаётся вопрос гарантии того, что из процесса корреляции не удаляются важные события. И, в-третьих, задача модуля анализа воздействия заключается в исключении из процесса корреляции сценариев атак, влияние которых на инфраструктуру сети

незначительно или невозможно, то есть в улучшении коэффициента редукции, при этом модуль анализа воздействия в блок редукции не входит.

В [46] предлагается разделение методов процесса корреляции событий безопасности по следующим уровням обработки данных (рисунок 8): (1) *первичные (сырые) данные* (*raw data*); (2) *события* (*events*); (3) *отчёты* (*reports*). В зависимости от конкретного уровня данные подвергаются соответствующей обработке для дальнейшего использования на более высоких уровнях, а в итоге - для принятия решения о возможных контрмерах и визуализации результатов.

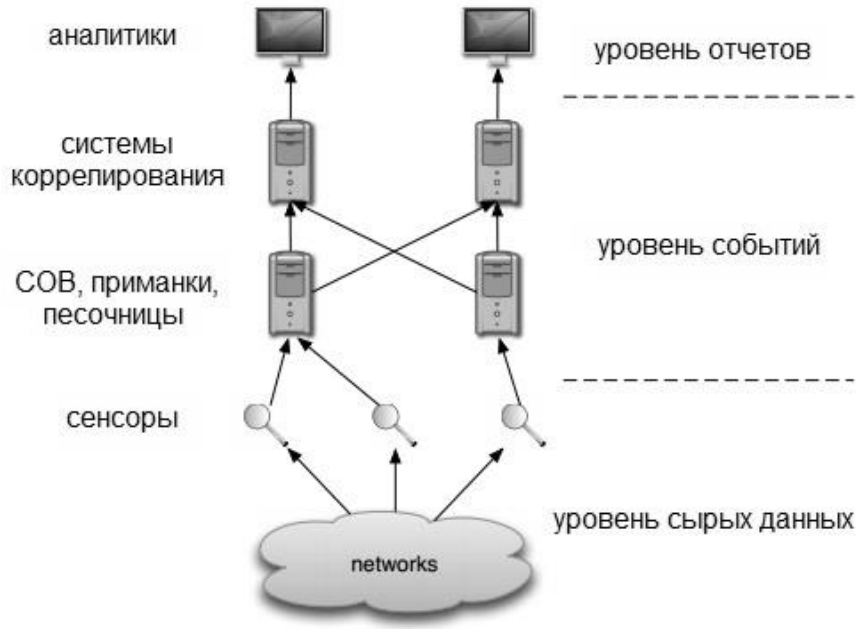


Рисунок 8 - Уровни обработки данных в процессе корреляции событий безопасности [46]

Уровень первичных данных состоит в основном из сетевых сенсоров, которые принимают необработанные данные и осуществляют их первичный анализ. Источниками данных для системы корреляции событий безопасности на данном уровне также могут быть системы журналирования различных сетевых приложений. На уровне первичных данных реализуются следующие процессы: *отбор пакетов* (*packet sampling*), *вероятностный анализ* (*probabilistic analysis*), *обнаружение аномальной активности* (*attack detection*), *обнаружение сканирования портов* (*detection of port scans*), *идентификация приложений*

(*application identification*) и *анализ полезной нагрузки пакетов* (*payload analysis*). На уровне событий выполняется локальный процесс корреляции событий безопасности и распределённый процесс корреляции событий безопасности, включая соблюдение конфиденциальности данных. На уровне отчётов специализированные приложения визуализируют результаты процесса корреляции наиболее подходящим образом. Уровень отчётов выполняет генерацию возможных активных контрмер и верификацию событий безопасности.

Отдельного внимания заслуживают работы, направленные на корреляцию данных с условно-статичным содержимым. Данные исследования представлены в [52], где были отмечены недостатки применения реляционных баз данных по сравнению с онтологическим подходом для целей мониторинга информации безопасности, когда необходимо иметь возможность логического вывода на основе данных в базе и простого перестроения модели данных. Онтологический подход в целом и способы и средства построения онтологий хорошо описаны в [53-55]. Работы, использующие онтологические модели для корреляции данных безопасности с условно-статичным содержимым, направленны на структуризацию и объединение данных безопасности. Как правило, данные работы касаются информации об уязвимостях, эксплойтах и программно-аппаратных продуктах, поскольку именно эти данные расширяются и используются намного активнее других. Так проект [56] предоставляет обобщенную информацию об уязвимостях и их реализациях в виде эксплойтов, поддерживает поиск информации по ключевым словам, названию продукта и многим другим. Работа [57] посвящена разработке интегрированной базы уязвимостей, включающей структурированную и обобщенную информацию об уязвимостях из разных источников, а также расширенный словарь продуктов.

В [58-60] представлены достаточно подробные онтологии для уязвимостей, однако в этих работах не уделяется должного внимания другим видам информации безопасности. В [61], напротив, подробно рассматриваются вопросы построения онтологического хранилища, однако не анализируются в должной

мере источники информации безопасности, которые могут быть использованы для его построения. Работа [62] сфокусирована на формировании онтологии показателей защищенности для применения при анализе безопасности и выборе контрмер, но не рассматривает подробно источники информации безопасности и вопросы формирования гибридного хранилища. Преимущества предлагаемого подхода заключаются в обеспечении эффективной интеграции данных в знания и поддержке получения новых знаний на основе данных из различных источников. Предлагаемый подход позволяет автоматизировать реализацию логического вывода при решении задач мониторинга и управления безопасностью [63].

Особо тщательное рассмотрение следует уделить модели Unified Cybersecurity Ontology (UCO), предложенной в [64,65]. UCO является онтологической моделью, интегрирующей различную информацию безопасности для оценки состояния защищенности в системе кибербезопасности. Данная онтология использует такие стандарты описания сущностей, как: CVE, CWE, CAPEC, CCE и др. Отличительная особенность данной модели заключается в наличии сущностей, описывающих информационно-коммуникационные объекты - файлы, сетевые адреса, процессы, операционные системы и др. Также довольно полезными являются сущности состояния сети и информация об атакующем. Однако UCO не позволяет интегрировать информацию из различных однотипных источников, поскольку модель не содержит соответствующих связывающих свойств. Поскольку часть концептов являются сугубо индивидуальными для мониторинга состояния безопасности конкретной инфраструктуры, используемое хранилище требует обязательной конфигурации при внедрении. Также сомнительными для использования являются сущности, отражающие текущее состояние системы в реальном времени, поскольку это требует оперативной модификации онтологической базы данных. Не смотря на большое количество баз данных безопасности, их отдельное применение не дает возможности получения общей картины о состоянии защищенности ввиду их значительной несогласованности. Совместное применение таких баз является

достаточно сложной задачей и требует больших временных затрат на предварительную обработку.

Этапы процесса корреляции. В рамках данной научно-квалификационной работы при дальнейшем рассмотрении процесса корреляции на основе работы [21] выделяются следующие его этапы:

Нормализация. Из-за того, что источники данных могут поставлять информацию в разном формате, возникает необходимость преобразования формата каждого события безопасности в некоторый нормализованный формат, который был бы понятен на всех последующих этапах обработки. Такое преобразование, или *нормализация*, означает, что синтаксис и семантика данных безопасности прозрачны и беспрепятственно определяемы системой корреляции.

Предобработка. После нормализации, обработанные данные безопасности нуждаются в дополнительной *предобработке*, так как часть источников может пропускать некоторые поля данных, важные для процесса корреляции (например, время начала, время окончания и источник данных).

Анонимизация. Данный этап процесса корреляции необходим, если производится работа с событиями минимум от двух источников, между которыми не установлено доверительное отношение. Анонимизация применяется для удаления или сокрытия конфиденциальной (или важной с юридической точки зрения) информации из данных безопасности. Существует две операции данного модуля: *анонимизация* и *псевдоанонимизация* [66,67]. Анонимизация препятствует восстановлению конфиденциальных данных, в то время как псевдоанонимизация является обратимой, а значит, оригинальные данные могут быть восстановлены доверенной стороной. В общем случае, желательно проводить псевдоанонимизацию, так как это позволяет получить доступ к оригинальной информации в ситуациях, когда необходим дальнейший анализ. Однако данное решение накладывает значительные вычислительные ограничения. Ключевой задачей модуля анонимизации является как сохранение необходимых свойств для анализа безопасности, так и способность их сокрытия от нежелательных сторон. Извлечение подобных свойств предполагает *деанонимизацию* (обратный процесс)

данных, что возможно только при использовании псевдоанонимизирующих методов.

Агрегация и фильтрация. Задача этапа фильтрации и агрегации заключается в удалении из потока входной информации системы корреляции данных по заранее определённым правилам (фильтрам), и в объединении данных, которые возникли в результате идентификации равнозначных данных. Решение о применении операции удаления принимается на основе определенных характеристик данных безопасности. Не прошедшие фильтрацию данные безопасности далее в процессе корреляции не участвуют, а оставшиеся - переходят на этап агрегации. Решение об агрегировании данных безопасности также принимается на основе их характеристик. При идентичности значений заранее определенной информации, а также удовлетворении временных характеристик событий заданному интервалу, такие данные безопасности агрегируют в более абстрактную высокоуровневую форму представления информации.

Восстановление хода атаки. Задача этапа восстановления хода атаки ограничена объединением данных безопасности, свидетельствующих об активности одного злоумышленника по отношению к одной цели (рисунок 9). Восстановление хода атаки построено на объединении данных безопасности с совпадающими атрибутами цели и источника атаки, временные характеристики которых попадают в заданный интервал. Использование временных характеристик заключается в окончания более раннего события, характерного для конкретной атаки, достаточно и определенно близко ко времени старта другого события, продолжающего соответствующую атаку.

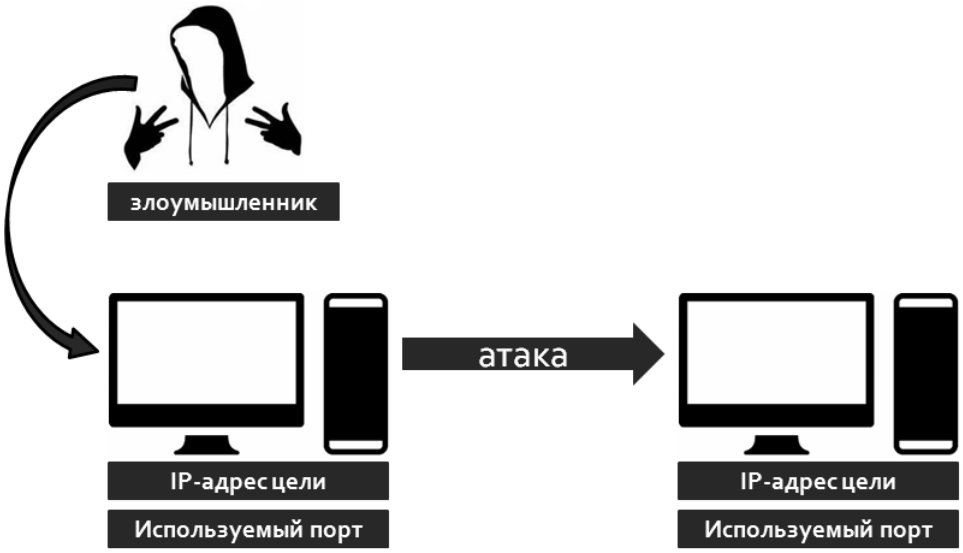


Рисунок 9. Восстановление хода атаки

Восстановление сессии атаки. Цель данного этапа – поиск связи между *сетевыми* (*network-based*) и *системными* (*host-based*) событиями безопасности (рисунок 10). Это необходимо для объединения ряда событий безопасности, вызванных злоумышленником при тестировании различных эксплоитов против определенной программы или запуском одного и того же эксплоита несколько раз для подбора правильных значений определенных параметров (например, смещений и адресов памяти для переполнения буфера). Процесс поиска связи между событиями усложняется за счет отличающегося предоставления информации в сетевых и системных событиях. Сетевые сенсоры могут предоставить информацию, характеризирующую обнаруженные атаки, например, IP-адреса источника и цели, используемые порты. Данные, поступающие от системных сенсоров, с другой стороны, содержат информацию об объекте, который был атакован, и субъекте, которым была осуществлена данная атака. Именно обнаружение связи между сетевыми и системными событиями позволяет определить отдельную сессию производимой атаки.

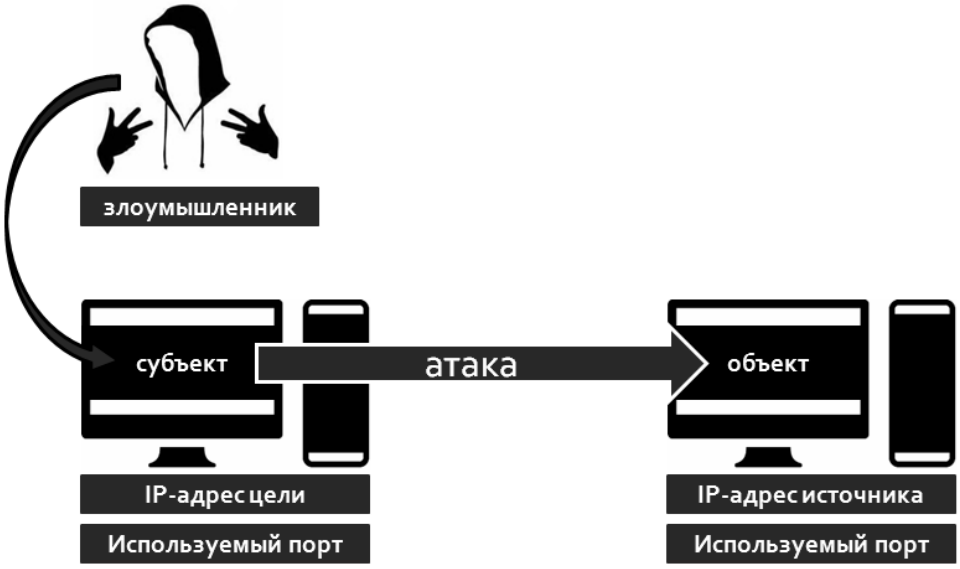
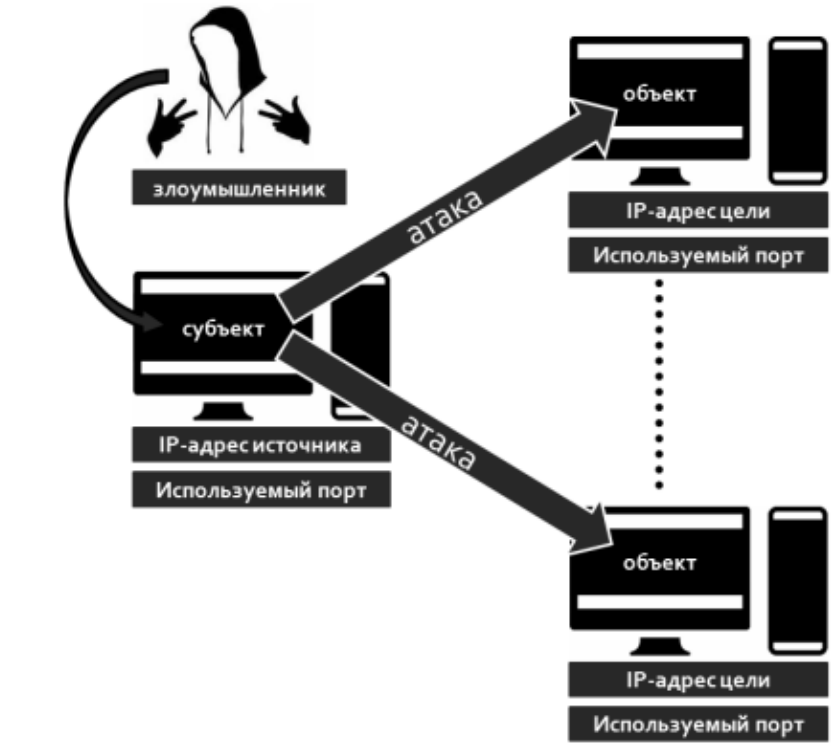


Рисунок 10 - Восстановление сессии атаки

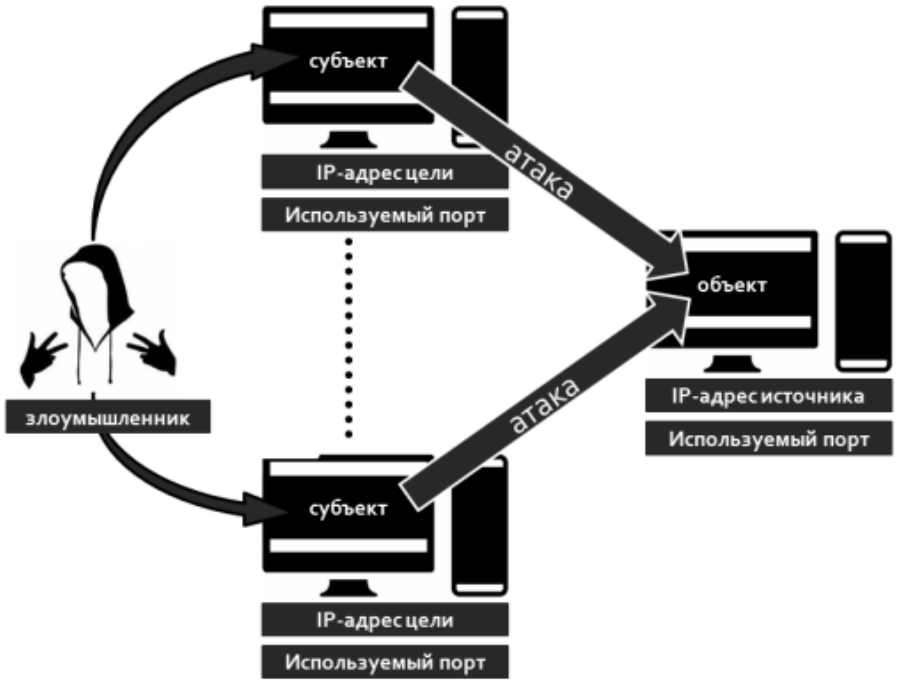
Определение цели и источника атаки. Задача данного этапа заключается в идентификации хостов (либо других элементов КФИ), которые являются либо источником, либо целью обнаруживаемых атак. Суть этапа заключается в объединении событий безопасности, ассоциируемых с отдельным хостом, который атакует несколько жертв (сценарий *один-ко-многим* представлен на рисунке 11,а), а также с несколькими хостами, которые атакуют одну жертву (сценарий *многие-к-одному* представлен на рисунке 11,б).

Многошаговая корреляция. Этап используется для распознания сложных сценариев, которые состоят из нескольких отдельных атак (рисунок 12). Обычно подобные сценарии определяются с использованием той или иной формы экспертных знаний [68-70]. Этап многошаговой корреляции также можно использовать для верификации высокоуровневых событий безопасности. При этом определяются сценарии атак, которые заведомо не имеют значения. Это позволяет удалить из процесса корреляции последовательности событий, которые коррелированны неверно. Например, при опросе сетевого окружения на предмет наличия какого-либо оборудования (принтера или сканера), последовательность событий по количеству запросов будет выглядеть как сканирование сети, однако, реально, данная последовательность не должна идентифицироваться как аномальная и фактически не является атакующим действием. К подобному роду

ошибок также можно отнести действия приложений, использующих в работе пиринговые (peer-to-peer, P2P) сети.



а) - Сценарий один-ко-многим



б) - Сценарий многие-к-одному

Рисунок 11 - Определение цели и источника атаки

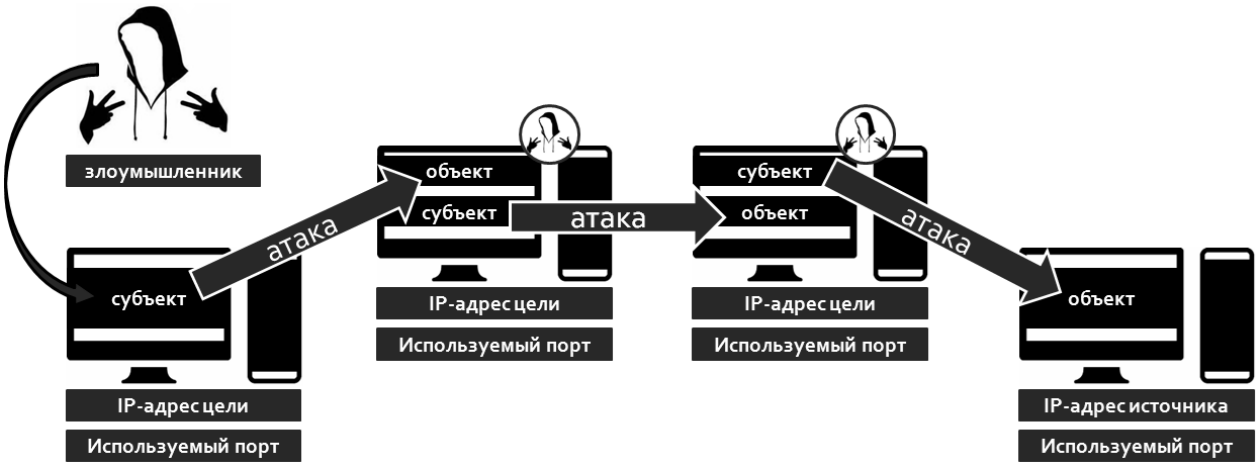


Рисунок 12 - Схема многошаговой корреляции

Анализ ущерба. Помимо аналитической информации, полученной на предыдущих этапах, данный этап использует стороннюю (не содержащуюся в событиях) информацию для анализа сценариев атак с точки зрения ущерба от их влияния на инфраструктуру сети или используемые ресурсы. На основе данных об ущербе, модуль назначает более высокую степень важности сценариям атак, которые угрожают более значимым активам сети. Информация о сети и соответствующих ресурсах хранится в хранилище активов, которая содержит подробности об используемых сетевых сервисах, зависимостях между ними, а также их важности для функционирования сети.

Приоритезация. Данный этапа должен учитывать политику безопасности и требования безопасности инфраструктуры, в которой развернута система корреляции. Фактически, этап ориентирован на проверку выполнения определенных условий, задаваемых администратором безопасности. Поэтому его основной задачей является выделение сценариев атак в соответствии с их приоритетом для задачи обеспечения безопасности.

Фильтрация на основе ранжирования. Этап используется для снижения количества рассматриваемых сценариев атак, критичность реализации которых мала для корректной работы защищаемой инфраструктуры. Данный модуль должен учитывать политику безопасности и требования безопасности инфраструктуры, в которой развернута система корреляции. Удаление из процесса корреляции сценариев атак с низким рангом (ущерб от влияния которых

на анализируемую сеть отсутствует или незначителен) снижает количество ошибок второго рода, увеличивая эффективность работы процесса корреляции.

Классификация методов корреляции. Для описания методов, используемых на разных уровнях и этапах обработки данных, необходимо дать определение понятия метода корреляции событий безопасности в рамках SIEM-систем. Метод корреляции включает последовательность действий над данными, направленную на выявление и (или) применение определенным способом признаков удаления, объединения, связывания, установления причинности и приоритетности обрабатываемой информации. Для удобства данные признаки можно обозначить как корреляционные признаки. Вместе с этим, существуют различные подходы, реализующие данные методы на этапах процесса корреляции.

В рамках процесса корреляции, от получения разноуровневых данных до формирования результатов, задачами методов корреляции являются [21,45,46,71]:

- преобразование данных для их понимания системой анализа защищенности компьютерной сети или SIEM-системой;
- преобразование данных от уровня к уровню для определения общего состояния анализируемой инфраструктуры;
- автоматизированное определение наиболее важных связей событий для их дальнейшего применения для анализа состояния инфраструктуры (самостоятельное обучение системы);
- приведение данных к виду, понятному администратору безопасности;
- сокращение данных до объема, приемлемого для обработки администратором безопасности.

Сами методы, применяемые в процессе корреляции, можно классифицировать в соответствии с моделью, представленной в таблице 1.

Параметр возможности изменения способа обработки корреляционных признаков отражает способность метода использовать операции различного характера для обработки корреляционных признаков. В данном случае, при

математическом описании методов корреляции следует обратить внимание на результирующую функцию каждого метода. Описываемая характеристика является показателем возможности изменения способа обработки аргументов данной функцией. Так, например, в методе на основе правил, обработка аргументов функции будет заключаться в сравнении их значений со значениями, заключенными в условиях. Таким образом, метод не подразумевает нескольких возможных способов обработки входных данных и поэтому является статичным по возможности изменения корреляционных признаков.

Таблица 1 – Классификационная модель методов корреляции

Параметр классификации метода корреляции	Класс метода корреляции
По возможности изменения способа обработки корреляционных признаков	Статические
	Динамические
По способу изменения корреляционных признаков	Самостоятельно изменяемые (самообучаемые)
	Изменяемые вручную
	Неизменяемые (необучаемые, фиксированные на начальном этапе)
По типу вычисления результата	Упорядоченные
	Вероятностные
	Смешанные
По возможности прослеживания пути вычисления результата	Определяемые
	Неопределяемые
По варианту оперирования корреляционными признаками	Использующие
	Определяющие
	Совместные

В свою очередь, нейронные сети, имеющие сумматорную и активационную функции, которые в конечно счете и являются результирующими, могут реализовывать различные способы обработки корреляционных признаков.

Способ изменения корреляционных признаков определяет методы по характеру изменения параметров обработки. Другими словами, опираясь на математическое описание метода как результирующей функции, данная характеристика определяет способ изменения результата в зависимости от

входных аргументов. Например, в методе на основе конечного автомата в качестве аргументов выступает множество входных состояний, определяемое заранее. Результат, принимаемый в данном случае, будет ограничен указанным множеством и графом переходов. Стоит отметить, что расширение множества входных состояний в общем случае приведет к перестроению графа переходов, что недопустимо во время работы метода (без остановки). Данный факт определяет метод как неизменяемый по выделенному признаку. В свою очередь, в правило-ориентированном методе аргументами функции являются сами правила, использующие множества текущих значений корреляционных признаков и сравнительных (эталонных) значений для принятия решения. Повлиять на результат в данном случае возможно при изменении самих правил, однако данный процесс не автоматизирован, что описывает метод как изменяемый вручную. Наконец, самообучаемые методы принимают в качестве аргументов значения корреляционных признаков, которые в данный момент могут влиять на результат работы при следующем использовании функции.

Тип вычисления результата характеризует методы с точки зрения используемого математического подхода. Данная характеристика описывает метод как упорядоченный при использовании в результирующей функции строгих логических структур. Например, в случае метода на основе конечного автомата и правило-ориентированного метода, выполнение каких-либо действий основано на строгом соответствии условия перехода и выполнения условия правила соответственно. Иными словами, в каждой точке результирующей функции по входным аргументам возможен только один вариант решения (множество решений упорядочено). Метод рассуждений на основе прецедентов, являясь адаптивным, может быть реализован как со строгим соответствием поведения инфраструктуры заданному значению, или с чисто вероятностным определением текущего поведения и выбираемого решения, так и со смешанным вариантом. Байесовские и нейронные сети по своей сущности основаны на вероятностных характеристиках корреляционных признаков. Данные методы используют веса (условные вероятности) влияния каждого корреляционного

признака на получаемый результат. Однако, для получения таких весов использование данных методов подразумевает наличие этапа обучения.

Возможность определения пути вычисления результата отражает способность обратимости алгоритмов, реализующих метод. Например, в случае метода на основе конечного автомата и правило-ориентированного метода, при известных входных и выходных параметрах путь выполнения алгоритма будет всегда однозначным и обратимо вычисляемым, то есть вычисляемым как от входных до выходных данных, так и наоборот. В случае с самообучаемыми методами корреляции, обратное преобразование от выходных к входным данным является трудновыполнимым или невыполнимым вовсе, так как входные данные уже могли повлиять на чувствительные элементы алгоритма (например, веса), а сохранение предыдущих состояний данных элементов не имеет особого смысла. Данная характеристика является полезной для определяемых методов при отладке и тестировании их конкретных реализаций, так как позволяет локализовать ошибку в работе алгоритма. В свою очередь, для неопределяемых методов, поиск ошибки обратным преобразованием выходных данных невыполним.

Параметр варианта оперирования корреляционными признаками дает представление о том, какую именно роль может выполнять метод в рамках процесса корреляции. Данное свойство выделяет использующие, определяющие и совместные методы. Другими совами, использующие методы описывают операции над признаками обработки для выполнения корреляции, а определяющие методы выполняют анализ данных для выявления признаков обработки для осуществления корреляции событий. Однако, существует ряд интеллектуальных методов, которые могут применяться как в роли определяющих, так и в роли использующих.

В ходе анализа работ, описывающих архитектуры, алгоритмы, методы и систем корреляции данных безопасности, было выделено пять методов, являющихся наиболее используемыми в области SIEM-систем, систем обнаружения и предотвращения вторжений и др. Данные методы отличаются друг от друга по ряду характеристик, учтенных в описанных параметрах

классификации. В свою очередь, данная разновидность позволяет использовать представленные методы в разных этапах процесса корреляции с учетом конкретных решаемых задач.

В таблице 2 представлены выделенные методы корреляции событий безопасности и их классификация в соответствии с рассмотренными признаками. Среди представленных в классификации методов не были добавлены такие методы, которые только определяют корреляционные признаки. Данное положение обусловлено ограничением применения подобных методов только в рамках обучения системы, что выходит за рамки данной статьи. К таким методам относятся методы кластеризации, построения деревьев решений, классификации, алгоритмы которых позволяют производить оценку качества выделенных корреляционных признаков, изначальное задание глубины анализа и многие другие характеристики.

В свою очередь, методы, являющиеся как использующими так и определяющими (то есть совместными) по оперированию с корреляционными признаками, применяются как на этапах обучения системы корреляции (определения признаков обработки), так и на этапах обработки данных процесса корреляции (применения признаков обработки). В литературе также рассмотрены такие методы корреляции, как кодовая книга [40,44], на основе сценариев/шаблонов атак [40,43], на основе модели состояния [44,45], голосование [45], явное изолирование ошибок [45], на основе графов зависимостей [45, 49], генетические алгоритмы [45, 71], школьная доска [45], на основе контекстно-независимой грамматики [45], на основе стандартного и аномального поведения [40,43], на основе иммунных систем [71], временно-ориентированные [43], на основе нечеткой логики [45,48,72], на основе схожести и др. [40,42-45,71].

Также используется ряд методов интеллектуального анализа событий для выявления корреляционных признаков. Стоит отметить, что несмотря на большое количество существующих методов корреляции, ряд методов могут быть логически преобразованы в другие методы. Например, метод на основе графа

зависимостей, метод конечных состояний и некоторые другие методы можно выразить правило-ориентированном методом.

Таблица 2 - Классификация методов корреляции

Метод корреляции	По возможности изменения способа обработки корреляционных признаков	По способу изменения корреляционных признаков	По типу вычисления результата	По возможности определения пути вычисления результата	По варианту оперирования корреляционными признаками
Конечный автомат	Статический	Неизменяемый	Упорядоченный	Определяемый	Использующий
Правило-ориентированный	Статический	Изменяемый вручную	Упорядоченный	Определяемый	Использующий
Рассуждение на основе прецедентов	Динамический	Самостоятельно изменяемый	Смешанный	Определяемый/ Неопределяемый (зависит от реализации)	Использующий
Байесовская сеть	Динамический	Самостоятельно изменяемый	Вероятностный	Определяемый/ Неопределяемый (зависит от реализации)	Совместный
Нейронная сеть	Динамический	Самостоятельно изменяемый	Вероятностный	Неопределяемый	Совместный

Ниже описаны принципы использования пяти выделенных методов обработки информации и проанализирована возможность их применения на различных этапах процесса корреляции.

Метод на основе машины конечных состояний (конечный автомат). Данный метод основан на построении модели графа переходов между состояниями анализируемой инфраструктуры. В качестве условий перехода выступают определенные параметры событий, тогда как само состояние определяет операции над анализируемым потоком событий. Метод конечных состояний включает в себя [45,71]: (1) множество возможных входных событий (входной алфавит);

(2) множество возможных выходных событий (выходной алфавит); (3) множество возможных состояний; (4) начальное состояние и (5) функцию перехода между состояниями.

В рамках корреляции событий безопасности, конечный автомат может быть как детерминированным, так и недетерминированным, поскольку в модели могут присутствовать как пустые (безусловные), так и двойственные переходы между состояниями, которые по единственному выполняемому условию ведут сразу к двум вершинам графа. В соответствии с классификацией, данный метод относится к статическим и неизменяемым поскольку построение модели графа переходов между состояниями производится сторонними средствами (не средствами самого метода) и до этапа эксплуатации системы. Однако, последующее изменение параметров модели (типов входных и выходных событий, множества состояний) возможно, но ведет к временной неработоспособности модулей, использующий данный метод.

Упорядоченность вычислений данного метода обеспечивается за счет того, что множества возможных входных и выходных событий, а также возможных состояний конечны. Также данный метод позволяет определить обратную цепочку вычислений за счет известности всех условий переходов между вершинами графа и при наличии конечного и начального состояний. Данное свойство является полезным при построении модели графа переходов. Метод конечных состояний применим в процессе корреляции только как использующий корреляционные признаки.

Описанный метод наиболее подходит для идентификации «вредных» (опасных, предупреждающих) состояний системы при мониторинге анализируемой инфраструктуры [45]. В рамках общего процесса корреляции реализация данного метода возможна на этапах агрегации и определения источника и цели атаки за счет предопределенных состояний. На этапах, использующих более сложные корреляционные признаки, метод конечных состояний применять затруднительно.

Правило-ориентированный метод. Данный метод является классическим и широко-распространенным не только в SIEM-системах, но и системах обнаружения и предотвращения вторжений, межсетевых экранах и антивирусных решениях. В основе данного метода лежат правила, имеющие понятный системе синтаксис и семантику [38,40,43-45]. Правило в данном методе является самостоятельной оперативной единицей (то есть операция может осуществиться за счет лишь одного правила). Каждое правило состоит из условия, проверяемого для входных данных по корреляционным признакам, и действия над поступившими данными, в случае выполнения условия. Все правила для каждой операции обработки данных находятся в хранилище правил. При поступлении данных на вход конкретной операции, они проходят проверку на предусловие, а именно, соответствие корреляционных признаков входной информации по отношению к правилам из хранилища. Также правила можно разделять на простые и сложные. Например, простым правилом можно назвать такую запись в таблице, для положительного исхода которой (применения правила) достаточно выполнения одного предусловия. В свою очередь, к сложным правилам относится набор из простых и сложных правил, связанных логическими операторами (И, ИЛИ, НЕ) и их комбинациями. Правила в хранилище можно добавлять, удалять и изменять в процессе работы всей системы. Однако изменение правил с помощью самих правил не предусмотрено, поэтому данный метод является статическим и изменяемым вручную. Последнее свойство также обусловлено сложностью составления самих правил. Правило-ориентированный метод является упорядоченным, определяемым и использующим корреляционные признаки, ввиду использования четкой логики выполнения правил, конечности их множества и отсутствия возможности применения для анализа данных с целью определения корреляционных признаков.

Правило-ориентированный метод применим в процессе корреляции на этапах нормализации, анонимизации, агрегации и фильтрации. Преимуществом описанного метода является строгое соблюдение условий при принятии решений, однако это не исключает логических ошибок, связанных с пересечением правил,

например, при удовлетворении предусловий сразу нескольких правил, с противоречивыми результатами их выполнения. Недостатком данного метода является большая емкость правил и сложность их составления для наиболее рациональной обработки как с точки зрения достижения оптимальной точности анализа, так и минимизации затрачиваемых ресурсов на его выполнение.

Метод рассуждений на основе прецедентов. В основу данного метода положена ситуационная модель, характеризующая поведение анализируемой инфраструктуры. Данная модель строится по обучающему множеству ситуаций (прецедентов), на основе которой определяется характер текущего поведения анализируемой инфраструктуры на этапе работы системы корреляции. Построение модели основано на использовании принципа адаптации. Данный принцип заключается в наполнении хранилища прецедентов (случаев) записями с возможными решениями. При поступлении нового прецедента, определяется наиболее подходящая запись из хранилища с соответствующим решением, после чего данное решение проходит проверку. Если подобного прецедента не существует, либо его решение не приемлемо, то система корреляции строит новое решение на основе старых. Полученное решение проверяется на корректность применения к прецеденту и, в случае успеха, вместе с ним добавляется в хранилище, иначе ищется новое решение [39,44,45].

Описанный метод является динамическим и самообучаемым, то есть неизвестные прецеденты, поступающие на вход системы, анализируются и добавляются в хранилище с наиболее подходящим решением. Метод на основе прецедентов по типу вычисления результата относится к смешанным, поскольку зависит от конкретной реализации, то есть может быть основан на упорядоченных и (или) статистических операциях. Ввиду наличия доступного хранилища также возможно определение пути по заданному конечному решению, однако при использовании вероятностных характеристик среди корреляционных признаков, поиск маршрута принятия решения может быть сильно затруднен либо полностью невозможен. Данный метод может использоваться в интеллектуальном анализе данных, но для определения новых корреляционных признаков в процессе

корреляции его применение невозможно, поскольку в данном случае необходимо полностью перестраивать ситуационную модель.

Метод на основе использования Байесовской сети. В основе данного метода лежит модель направленного ациклического графа. Суть метода заключается в расположении в вершинах сети корреляционных признаков, а связывающие их направленные дуги задают отношения условной независимости их значений [39,40,43,45]. Обучение системы корреляции производится за счет последовательного вычисления значений условных вероятностей вершин, переменные которых неизвестны. Данная операция выполнима за счет подачи на вход данных со значениями переменных корреляционных признаков. Метод подразумевает наличие обучающей выборки событий безопасности и является динамическим, самообучаемым и вероятностным. Путь прохождения результата вычисления на этапе работы Байесовской сети определяется только в случае сохранения множества корреляционных параметров в вершинах сети [45].

Для определения корреляционных признаков, то есть таких признаков, которые влияют на установление наличия связи между событиями и их причинно-следственные отношения, в вершинах графа размещаются признаки событий. Так же, как и обучение системы корреляции, поиск корреляционных признаков требует обучающей выборки. В случае продолжения обучения (коррекции) системы на этапе работы, Байесовская сеть будет динамической. Другими словами, определение корреляционных признаков и корреляция событий будет происходить одновременно в рамках одной модели. Исходя из данного свойства, описанный метод по вариантам оперирования корреляционными признаками является смешанным. Однако, в таком случае, определение пути получения результата будет затруднено или невозможно.

Представляется, что в общем процессе корреляции за счет динамического и самостоятельного обучения метод на основе использования Байесовской сети наилучшим образом может быть применен на этапах многошаговой корреляции, анализа ущерба и приоритезации.

Метод на основе использования искусственной нейронной сети. Основой данного метода является математическая модель, состоящая из нейронов, имеющих собственное состояние, и линий связи (синапсов), определяющих влияние входных для него нейронов на данное состояние. Результатом выполнения работы каждого нейрона является аксон, значение которого может быть использовано в качестве входных нейронов для нейронов более высокого уровня [39,40,45,48,71]. В рамках корреляции событий безопасности в роли нейронов выступают корреляционные признаки. В любой схеме нейронной сети содержатся минимум 2 уровня - нулевой и единичный. На этапе обучения на входные нейроны поступают множества значений корреляционных признаков потока событий, при этом влияние входных нейронов на нейроны следующего уровня изначально задается случайно [45,48]. По мере обучения системы, значения влияний корректируются для соответствия заданному результату. Такой подход является обучением с учителем.

На этапе работы системы корреляции с использованием данного метода вычисленные значения влияний могут корректироваться, поэтому данный метод относится к динамическим и самообучаемым.

Как и в случае с использованием Байесовской сети, значения влияний задаются вероятностным отношением - следовательно, метод является вероятностным. Метод также позволяет анализировать входные данные и корректировать множество корреляционных признаков в процессе выполнения. Однако даже если система корреляции не продолжает обучение нейронной сети в процессе выполнения, вычислить обратный путь следования от результата не представляется возможным.

Метод на основе использования искусственной нейронной сети, также как и предыдущий, может быть использован на этапах многошаговой корреляции, анализа ущерба и приоритезации. В то же время, оба метода можно применять для определения корреляционных признаков в рамках методов, которые не могут это сделать самостоятельно.

Комбинированные (гибридные) и другие методы корреляции. В реальных условиях применение только одного метода корреляции для анализа исследуемых инфраструктур недостаточно для получения точной оценки защищенности компьютерной сети и управления событиями и информацией безопасности [39,45,71]. Данное положение обусловлено рядом факторов, таких как: (1) вычислительная сложность метода; (2) функциональные возможности метода; (3) ресурсопотребление и др. Используя несколько методов на критичных этапах процесса корреляции, при пересечении множеств получаемых результатов возможно добиться более высокой точности оценки защищенности анализируемой инфраструктуры и определения текущей ситуации по компьютерной безопасности. Также возможен вариант последовательного применения разных методов корреляции на одном из этапов общего процесса корреляции. Данный факт связан с обработкой данных разных уровней, например при анализе простых и более сложных событий безопасности.

В работе освящены далеко не все имеющиеся на данный момент методы корреляции событий безопасности, но наиболее востребованные с технической точки зрения в рамках представленной схемы процесса корреляции.

1.4 Требования, предъявляемые к системе корреляции данных для мониторинга и управления безопасностью в КФС

По результатам анализа работ в предметной области были сформулированы требования к методикам корреляции информации безопасности, в основу реализации которых должен быть положен модельно-методический аппарат, разрабатываемый в данном исследовании. Требования условно разделяются на функциональные и нефункциональные. Функциональные требования определяют функционал, который должна выполнять система, реализующая разрабатываемые модели и методики методики. Нефункциональные требования представлены рядом ограничений, налагаемых на потребляемые системой ресурсы (например, временные ограничения, перечень используемых стандартов) [73].

Для достижения основной цели исследований, разрабатываемая система корреляции должна реализовать минимум 2 вида процедур: (1) процедуры адаптации системы к ЦИ КФС, включающие автоматизированное выявление информационных объектов, их типов, характера взаимодействия, семантических связей и зависимостей, сценариев поведения, а также определение правил предварительной обработки информации, за счет анализа исторических и архивных данных безопасности; (2) процедуры мониторинга текущего состояния защищенности, включающего определение текущего поведения информационных объектов и его характера, а также прогнозирование вероятности происхождения того или иного события (инцидента) безопасности. Стоит отметить, что процедуры первого вида могут быть выполнены до непосредственного внедрения системы корреляции в ЦИ КФС в так называемом режиме «off-line». Однако, данное положение не исключает возможности выполнения подобных процедур исключительно в реальном масштабе времени. В свою очередь, процедуры второго вида могут быть выполнены только в реальном масштабе времени, то есть – «on-line». Однако преодоление реального размещения системы корреляции в ЦИ КФС возможно за счет имитационного моделирования ее работы в течении времени после выполнения адаптационных процедур. Таким образом, разрабатываемая система должна осуществлять обработку информации в контексте двух основных направлений: (1) адаптации к ЦИ КФС и (2) проактивном мониторинге состояния. Очевидно, что выполнение первого направления является первоочередным, тогда как выполнение второго направления возможно только на основе результатов работы первого.

Функциональные требования к системе корреляции, в которой реализованы разрабатываемые модели и методика, были разделены на общие требования, требования, предъявляемые к процедурам адаптации к ЦИ, и требования, предъявляемые к процедурам проактивного мониторинга состояния защищенности КФС. Общие требования:

6) Система должна учитывать разнородность исходных данных безопасности.

7) Система должна поддерживать обработку частично анонимизированных (псевдо-ананимизированных) исходных данных.

8) Система должна учитывать неопределенности переменного (типового), конфигурационного и концептуального уровня ЦИ КФС.

9) Система должна обеспечивать обработку больших массивов исходных данных безопасности.

Функциональные требования, предъявляемые к системе при реализации проактивного мониторинга состояния защищенности НЦИ КФС:

1) Система должна выполнять анализ поведения активов и взаимодействующих с ними объектов НЦИ КФС.
2) Система должна формировать модели популяций активов и взаимодействующих с ними объектов НЦИ КФС.
3) Система должна прогнозировать вероятность возникновения конкретного события (предупреждения, инцидента) в будущем отрезке времени.

Нефункциональными требованиями к системе корреляции данных безопасности являются ограничения предоставляемых ресурсов на ее функционирование: (1) по времени с точки зрения оперативности; (2) по аппаратным ресурсам с точки зрения ресурсопотребления; а также поддержка (3) масштабируемости обработки информации за счет параллельного выполнения отдельных операций комплексной методики корреляции данных безопасности.

1.5 Постановка задачи исследования

Система корреляции данных безопасности должна обладать способностью адаптироваться к НЦИ КФС в условиях условно-неограниченного количества источников информации. Для выполнения выдвинутых требований к системе корреляции предполагается проведение статистического, структурного и интервального анализа исходной информации и построение на основе их результатов модели ЦИ КФС. Данная модель должна обеспечивать связь

статического состояния защищенности КФИ с его динамическим изменением в течении времени для формирования общей модели поведения реальных информационных объектов ЦИ. Предполагается, что подобный подход позволит строить предсказательные модели наступления конкретных событий (предупреждений, инцидентов) для определенных информационных объектов.

Результаты работы системы корреляции по направлению адаптации заключается в: автоматизированном определении модели НЦИ КФС и построение модели поведения ее информационных объектов. В свою очередь, реализация направления проактивного мониторинга состояния защищенности ЦИ КФС в каждый момент времени должна формировать вероятность наступления конкретных инцидентов.

Задачу разработки системы корреляции данных безопасности следует разделить на подзадачи:

- Определение модели НЦИ КФС с учетом предопределенных видов и типов входной информации.
- Разработка модели корреляции данных с условно-статичным содержимым как между типами входной информации, так и среди разнородных источников одного типа.
- Разработка модели корреляции данных с динамичным содержимым.
- Разработка комплексной методики корреляции данных безопасности с возможностью автоматизированной адаптации к НЦИ КФС.

На содержательном уровне научную задачу научного исследования можно сформулировать следующим образом: разработать модельно-методический аппарат (комплекс моделей, методик и алгоритмов), реализующий адаптивную корреляцию данных безопасности для оценки защищенности КФС. Реализация таких моделей, методик и алгоритмов в системах оценки защищенности (например, SIEM-системы следующего поколения) должны существенно снизить трудоемкость процесса конфигурации подсистемы корреляции и повысить точность и оперативность оценки состояния защищенности на основе анализа эмпирических данных безопасности.

Выводы по главе 1

1) Проведенный анализ работ в области корреляции данных безопасности показал отсутствие подхода, осуществляющего адаптивную поддержку НЦИ и изменяющихся ЦИ, тогда как данная задача является важной и актуальной с учетом возрастающей сложности инфраструктур различных КФС.

2) Определены место и роль процесса корреляции в SIEM-системе. Описаны основные этапы корреляции, их роль и необходимость использования, а также произведен их анализ с точки зрения реализации системы оценки защищенности компьютерных инфраструктур.

3) Разработана классификация методов корреляции, что позволяет выделить преимущества и недостатки каждого из методов для наиболее эффективного применения на разных этапах процесса корреляции.

4) Определены функциональные и нефункциональные требования к системе корреляции данных безопасности.

5) Описана постановка задачи исследования и определены основные направления работы для достижения цели научно-квалификационной работы.

Глава 2 Модели и комплексная методика корреляции больших массивов гетерогенных данных для мониторинга и управления безопасностью в КФС

2.1 Модель неопределенной инфраструктуры КФС

Пусть в каждый момент времени, неопределенная инфраструктура *I* состоит из множества информационных объектов *O*:

$$O^I = \{o_{1,} o_2, \dots, o_s\}, \quad (1)$$

существующих во времени (то есть имеющих какую-то продолжительность жизни), и состояние которых описывается с помощью одной или нескольких характеристик *x* из множества характеристик *X*:

$$o = \{x_1 : v^{x_1},\ x_2 : v^{x_2}, \dots,\ x_k : v^{x_k}\},\ o\ \in O^I,$$
$$x\ \in X\ , X = \{x_1, x_2, \dots, x_r\},\ |X| \geq 1, \quad (2)$$

где s и r – количество объектов и общего числа их характеристик соответственно. Информационные объекты обязательно связаны друг с другом отношениями принадлежности. То есть каждый объект обязательно является частью более высокоуровневого объекта и (или) содержит в себе более низкоуровневые объекты. Также связь между объектами определяется за счет их непосредственного взаимодействия друг с другом.

Предполагается, что набор характеристик X^o, которыми описывается информационный объект *o*, однозначно определяет тип информационного объекта *ot*:

$$X^o = ot,\ o \in O,\ X^o \subset X,\ ot \in OT,\ OT = \{ot_1, ot_2, \dots, ot_m\}, \quad (3)$$

где m – количество типов информационных объектов множества *OT*. При этом каждая характеристика *x* из множества *X* принадлежит только одному типу информационного объекта *ot*:

$$ot = X^{ot},\ X^{ot} \subset X,\ X^{ot_i} \cap X^{ot_j} = \emptyset\ \ \forall\, i, j\ \in \{1,2, \dots, m\}, i \neq j \quad (4)$$

Заключительным элементом модели неопределенной инфраструктуры I является множество отношений R между объектами O. Исследование данного множества является дальнейшим направлением работы.

Таким образом, модель M неопределенной инфраструктуры I состоит из множества информационных объектов O, а также множеств их типов OT и отношений OR:

$$M^I = < O, OT, OR > \quad (5)$$

Стоит отметить, что модель НЦИ определяется поведением элементов системы, которое отражается в данных с динамичным содержимым. Однако данные с условно-статичным содержимым в данной модели выражаются характеристиками объектов.

2.2 Модель корреляции информации с условно-статичным содержимым на основе онтологического подхода

В ходе анализа источников данных, участвующих в построении гибридного хранилища информации безопасности, были получены статистические характеристики связей между различными базами (см. 1.2), представленные в таблице 3. В данной таблице отображены следующие показатели: (1) наименование релевантной базы (B2), на которые имеются ссылки в анализируемой (целевой) базе (B1); (2) общее количество ссылок B2 в B1; (3) количество записей B1, имеющих ссылки на B2; (4) количество уникальных элементов из B2 в B1; (5) процент элементов B1, в которых есть ссылки на B2; (6) процент уникальности используемых ссылок B2 в B1; (7) среднее количество ссылок B2 на один элемент B1; (8) используемость базы B2 в базе B1. В данном случае B1 обозначает целевую базу данных, а B2 – стороннюю связанную базу. Стоит также отметить, что указание ссылок анализируемых баз на самих себя (NVD-CVE, CAPEC,EXPLOIT-DB) означает общее количество записей в данных базах, например база «CAPEC» имеет 528 записей.

В результате анализа описанных в разделе 1.2 типов исходных данных с условно-статичным содержимым, а также конкретных источников информации безопасности, был выявлен наиболее распространенный недостаток существующих баз: сильная разобщенность данных и несогласованность их форматов в значительной степени препятствует формированию общей картины связанных данных и их влияния на аспекты безопасности КФС.

Таблица 4 - Статистические характеристики связей между источниками информации безопасности с условно-статичным содержимым для концептов онтологии

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
NVD							
CVE	84557						
EXPLOITS	2752	2599	2409	3,1%	87,5%	0,03	7,21%
CPE	2048178	82740	199160	97,9%	9,7%	24,22	169,97%
CWE	51131	50519	88	59,7%	0,2%	0,60	12,45%
CAPEC							
CAPEC	528						
CVE	69	36	57	6,8%	82,6%	0,13	0,07%
CWE	964	234	241	44,3%	25,0%	1,83	34,09%
EXPLOIT-DB							
EXPLOIT-DB	33394						
CVE	33993	33993	16879	100,0%	49,7%	0,99	19,96%

Онтологический подход является одним из решений для представления взаимосвязанных данных с учетом сложности их структур. Это позволяет выражать сложные отношения между объектами, использующими логику описания. Подход состоит в определении набора понятий в выбранной предметной области. Тогда связи между понятиями генерируются с учетом их отношений и взаимодействия.

Сущности онтологической модели определяются с учетом следующих информационных объектов: уязвимости, ПАО, слабости, эксплойты, шаблоны атак, безопасные конфигурации и контрмеры. Чтобы идентифицировать

гибридную структуру взаимосвязей между этими объектами, мы рассмотрели основные открытые базы данных и обозначили отношения между ними. На рисунке 13 представлена иерархия наследования классов онтологической модели для гибридного хранилища информации безопасности. В качестве корневых сущностей выступают 7 типов информации безопасности с условно-статичным содержимым (выделены оранжевым), а также сущность «Источники», которая выступает для более явного разделения как разнотипных, так и однотипных источников информации безопасности. Стрелками в данной схеме обозначена принадлежность сущности-потомка к сущности-предка (родителя), например сущность «CVE» является потомком сущностей «Уязвимости» и «Источники».

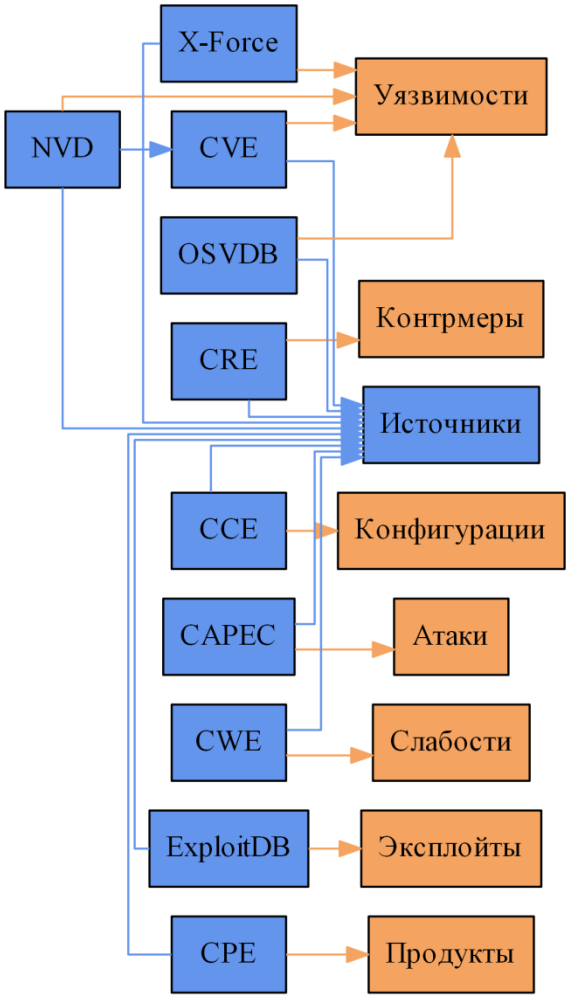


Рисунок 13. Иерархия наследования классов онтологической модели гибридного хранилища информации безопасности

Все классы условно разделяются на: (1) типы информации безопасности (все кроме классов «Источники» и «Ссылки» и их наследников); (2) источники

информации безопасности («Источники» и дочерние классы); (3) ссылки на сторонние источники описания конкретной информации безопасности («Ссылки», на рисунке 13 не указаны). Представленная модель описывает не только родительские отношения между понятиями области знаний «Информация безопасности», но также учитывает отношения принадлежности между концептами (типами) и конкретными источниками данных. Однако иерархия наследования классов не отражает связи между сущностями(свойства-объекты) и возможных вариантов описания их индивидов (свойства-значения).

В предыдущих работах [24] было установлено, что класс уязвимостей имеет наибольшее количество связей с другими классами информации безопасности в онтологической модели. Поэтому, определение объектных свойств будет начинаться непосредственно с данного класса. Выделено 4 иррациональных свойства: (1) «реализуетсяПосредством»; (2) «реализуетсяВ», (3) «реализуетсяПри», (4) «реализует». Иррациональность представленных свойств-объектов заключается в ограничении экземпляров класса уязвимостей использовать данные отношения к самим себе. Иными словами, два индивида, связываемых друг с другом посредством перечисленных свойств, должны быть обязательно разными.

На основе указанных объектных свойств справедливы следующие аксиомы для сущности «Уязвимость»: (1) «реализуетсяПосредством Эксплойта»; (2) «реализуетсяВ Продукте»; (3) «реализуетсяПри УязвимойКонфигурации»; (4) «реализует Слабость»; (5) «реализует Атаку». Таким образом, перечисленные свойства характеризуют отношения уязвимостей к 5 другими видам информации. Полное описание отношений между сущностями представлен на рисунке 14.

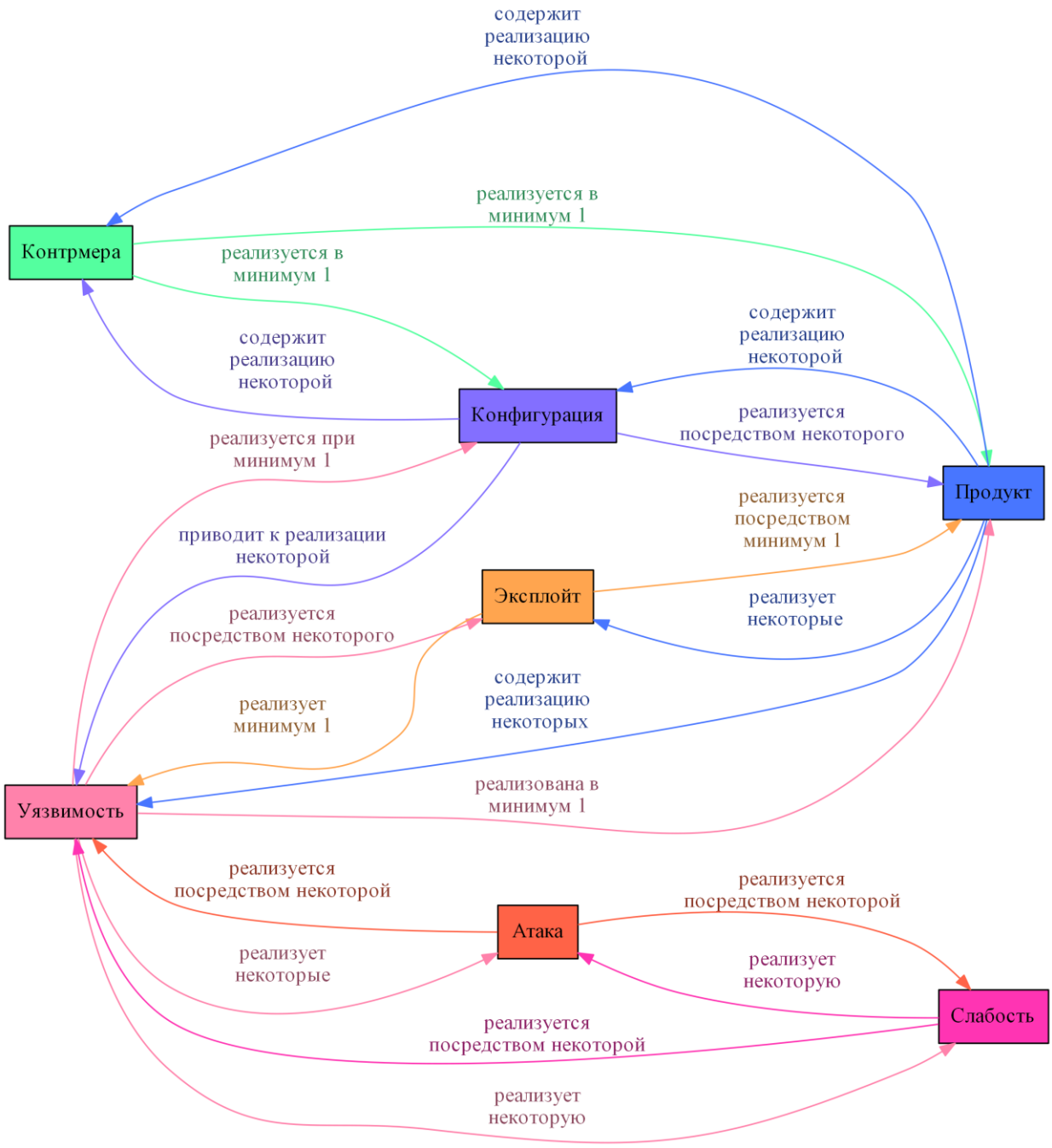


Рисунок 14 – Отношения между сущностями онтологической модели корреляции данных со статичным содержимым

Однако также существуют отношения между однотипными и разнотипными экземплярами классов информации безопасности, не связывающие концепты между собой напрямую (посредством прямого отношения). Данное объектное свойство характеризует связь субъекта с каким-либо внешним источников (например, по ссылке на URL или иному условно-уникальному идентификатору), выраженной сущностью «Ссылка». Таким образом, содержащиеся в гибридном

хранилище индивиды обладают объектным свойством «имеетСвязьС». В качестве объекта данного утверждения и диапазона допустимых значений выступают экземпляры класса «Ссылка». Указанное свойство является симметричным, поскольку связь индивида с внешним источником равнозначна связи внешнего источника с индивидом. Подобное утверждение позволяет связывать с помощью сущности «Ссылка» как однотипные, так и разнотипные индивиды гибридного хранилища. Общее перечисление свойств-объектов в разработанной онтологической модели, а также их специфических характеристик представлено в таблице 4.

Если объектное свойство является транзитивным и связывает индивида *a* и индивида *b*, а также индивида *b* связывает с индивидом *c*, то можно заключить, что индивид *a* связан с индивидом *c* через данное свойство. Например, для аксиом: «Уязвимость реализуетсяПосредством Эксплойта» и «Слабость реализуетсяПосредством Уязвимости» справедливо утверждение: «Слабость реализуетсяПосредством Эксплойта». Описание характеристик иррациональности и симметричности приведено выше.

Таблица 4. Объектные свойства онтологической модели гибридного хранилища информации безопасности

Объектное свойство	Обратное свойство	Транзитивное	Симметричное	Иррациональное
реализуется Посредством	реализует	+	-	+
реализует	реализуется Посредством	+	-	+
реализуетсяВ	содержит Реализацию	+	-	+
содержит Реализацию	реализуетсяВ	+	-	+
реализуетсяПри	приводитК Реализации	-	-	+
приводитК Реализации	реализуетсяПри	-	-	+
имеетСвязьС	-	+	+	+

Стоит отметить, что в представленной таблице заполнены только те ячейки, концепты которых непосредственно связаны объектным свойством. Иными словами, связи, вычисляемые на основе логического вывода за счет транзитивности и симметричности свойств, не рассматриваются. Также, при разработке онтологической модели гибридного хранилища, отношения между объектами задавались на основе не только теоретического представления природы концептов, но и практического использования источников информации безопасности.

С учетом применения ограничения кардинальности множества объектов для конкретных свойств, и при условии, что каждый верхнеуровневый класс информации безопасности непосредственно (явно) относится к другим верхнеуровневым классам посредством только одной пары объектных свойств (прямого и обратного), общая картина связанности концептов онтологической модели представлена в таблице 5.

Часть сущностей, характеризующих индивидов в онтологической модели реализовано за счет определения свойств данных. Подобными свойствами являются поля описания информации безопасности. Например, определение иерархии метрики CVSSv2 для описания уязвимостей представлена на рисунке 15 (слева). В данном случае, важной особенностью языка OWL2 является возможность задания области значений для свойства данных на основе пользовательского типа. Например, диапазон возможных значений для метрики «AccessComplexity» задается выражением (в редакторе Protege 5.0):

{«*High*», «*Low*», «*Medium*»}.

На рисунке 15 (справа) представлен пример определения иерархии описания для концепта «Продукт» («Product»).

Язык OWL2 также позволяет задавать область определения как для объектных свойств, так и для свойств данных, в виде одного или нескольких классов (доменов). Для указанных выше примеров определений метрики CVSSv2 и записи CPEv2 в качестве доменов заданы классы «Уязвимость» и «Продукт» соответственно.

Таблица 5. Типы отношений между основными концептами онтологической модели гибридного хранилища информации безопасности.

	Продукт	Уязвимость	Атака	Слабость	Эксплойт	Конфигура-ция
Продукт	-	содержит реализацию некоторых	-	-	реализует некоторые	-
Уязви-мость	реализуется в минимум 1	-	реализует некоторые	реализует максимум 1	реализуется посредством некоторых	реализуется при минимум 1
Атака	-	реализуется посредством некоторых	-	реализуется посредством некоторых	-	-
Слабость	-	реализуется посредством некоторых	реализует некоторые	-	-	-
Конфигу-рация	реализуется посредством минимум 1	-	-	-	-	-
Контрмера	реализуется в минимум 1	-	-	-	-	реализуетсяВ некоторых
Эксплойт	реализуется посредст-вом минимум 1	реализует минимум 1	-	-	-	-

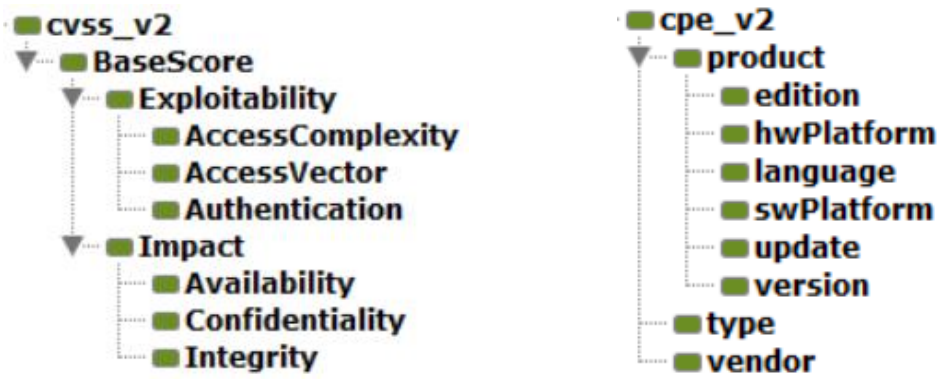


Рисунок 15. Определение свойств данных в онтологической модели гибридного хранилища.

Также мощным инструментом онтологического подхода является возможность задания эквивалентности между сущностями. Иными словами, характеристика эквивалентности может быть использована для определения отношений между классами, объектными свойствами и свойствами данных. Например, текстовое описание уязвимости в базе NVD представлено в поле «Descrption», тогда как в базе X-Force эквивалентным смыслом обладает поле «Overview», а в условно-неизвестной базе – поле «Annotation». Таким образом, при задании отношения эквивалентности между приведенными свойствами данных и дальнейшем анализе индивидов из разнородных баз в рамках онтологической модели можно оперировать любым из указанных понятий.

Стоит также раскрыть возможность описания количественных ограничений (кардинальности) множеств используемых объектов для объектных свойств. Данный механизм подразумевает присвоение значения использования конкретного свойства индивидом класса и (или) количество индивидов, связываемых посредством данного свойства. Например, утверждения «Каждый Эксплойт реализует минимум 1 Уязвимость» и «Каждая Уязвимость реализуетсяВ минимум 1 Продукте» высказываются о каждом субъекте («Эксплойт», «Уязвимость»), который связан минимум с одним объектом («Уязвимость», «Продукт») посредством соответствующих свойств. В действительности, индивид класса «Эксплойт» не может существовать без индивида класса «Уязвимость» (для второго утверждения справедлива аналогичная аксиома). Однако ввиду большой несогласованности источников информации безопасности, характеристика кардинальности свойств была использована с осторожностью. Например, логично предположить, что любая уязвимость реализует какую-либо слабость. Однако далеко не все уязвимости можно связать с базой слабостей, поскольку в исследуемых источниках нет соответствующих прямых и перекрестных ссылок (идентификаторов). В данном случае, было использовано определение кардинальности с помощью следующего утверждения: «Некоторая Уязвимость реализует максимум 1 Слабость». Данная конструкция является более гибкой, чем фиксированное ограничение кардинальности и позволяет исключить

ошибки недостаточности описания индивида при использовании онтологической модели гибридного хранилища.

2.3 Модель корреляции данных с динамичным содержимым

Для определения модели корреляции данных с динамичным содержимым сначала требуется задать формальное описание входной информации. Пусть исходными данными для выполнения процесса корреляции анализа является множество событий безопасности *E* журнала *L*.

$$E^L = \{e_1, e_2, \dots, e_n\} \quad (6)$$

Под событием понимается факт либо результат происхождения какого-либо действия на любом из его этапов: попытка (событие отказа), начало действия (событие старта), промежуточный результат действия (события, продолжительные по времени), конечный результат выполнения действия (завершено корректно, завершено с ошибкой). Каждое событие e состоит из множества свойств p и их значений v, которыми описывается определенное действие:

$$e = \{p: v\}, P = \{p_1, p_2, \dots, p_d\}, V = \{V^{p_1}, V^{p_2}, \dots, V^{p_d}\}, \quad (7)$$

где P и V^p - множества свойств и множества их значений соответственно, а d – общее количество свойств. В каждый момент времени данные множества являются конечными и не пустыми, хотя в течение времени они могут изменять свою мощность:

$$P \neq \emptyset, V^{p_i} \neq \emptyset, \forall\, i \in \{1,2, \dots, d\} \quad (8)$$

Соответственно, каждое свойство p события e определено значением v из множества возможных значений V^p:

$$p^e = v, v \in V^p, V^p = \{v_1, v_2, \dots, v_h\}, V^p \in V, \quad (9)$$

где h – количество возможных значений свойства p.

Дальнейшее описание модели корреляции данных с динамичным содержимым основано на модели неопределенной инфраструктуры. Допускается,

что изменения состояния целевой инфраструктуры *I* описывается событиями *E* в журнале *L*. Связь между событиями безопасности *E* и объектами инфраструктуры *O* строится на следующем утверждении: каждое свойство *p* события *e* является характеристикой *x* объекта *o*:

$$p^e = x^o \tag{10}$$

Другими словами, каждое событие описывает минимум один информационный объект инфраструктуры.

Исходя из выражения (7), каждое событие *e* имеет определенный набор свойств P^e, с помощью которых описываются действия:

$$e = P^e, P^e \subset P \tag{11}$$

Предполагается, что различные действия, производимые множеством объектов *O*, описываются различными типами *ET* событий *E*. Другими словами, каждый тип *et* является уникальным (неповторяющимся среди других типов) сочетанием свойств P^{et} :

$$ET = \{et_1, et_2, \dots, et_q\}, et = P^{et},$$
$$P^{et} \neq \emptyset,\ P^{et} \subset P, P^{et_i} \neq P^{et_j}\ \forall\ i, j\ \in \{1, 2, \dots, q\}, \tag{12}$$

где q – количество типов событий.

Исходя из вышеописанного, формальная постановка задачи адаптации процесса корреляции данных безопасности с динамичным содержимым заключается в разработке подхода *f*, который направлен на преобразование исходных данных к множествам объектов и их типов модели неопределенной инфраструктуры:

$$\{E, P, V\} \xrightarrow{f} \{O, OT\} \tag{13}$$

2.4 Комплексная методика корреляции гетерогенных данных в киберфизических системах

Первоначальным этапом выполнения комплексной методики корреляции является структурный анализ событий безопасности, основанный на

исследовании их (событий) свойств, а также связей между свойствами. Предлагается поделить множество видов отношений между свойствами на (1) однотипные и (2) разнотипные. Однотипные свойства могут быть эквивалентными по 2 признакам: (1) абсолютного сходства и (2) семантической (типовой) схожести. Разнотипные свойства могут быть эквивалентными по различным признакам: уникальности, равновероятности, используемости и др. Каждый вид отношений между свойствами событий безопасности формирует множество соответствующих видов связей, которые обозначаются символом ~ с указанием признака эквивалентности вида отношения:

$$p_1 \overset{abs}{\sim} p_2, \tag{14}$$

что означает связь свойства p_1 с свойством p_2 по признаку абсолютной равнозначности, или свойство p_1 абсолютно эквивалентно свойству p_2. Таким образом, определенные виды отношений между свойствами задают соответствующие характеристики свойств относительно друг друга. Функции определения связей между любыми двумя свойствами p_1 и p_2 обозначаются $f^{feat}()$, где *feat* – признак эквивалентности вида отношения, реализующего связь между свойствами.

Далее, каждый исследованный в данной работе вид отношений между свойствами для проведения структурного анализа будет рассмотрен отдельно.

Однотипные отношения по признаку абсолютного сходства между свойствами событий e_1 и e_2 определяют схожесть их типов, если каждое свойство события e_1 имеет данную связь со свойством e_2, и наоборот:

$$et^{e_1} = et^{e_2} \Leftrightarrow \forall\, p^{e_1}, p^{e_2} \in P \;:\; \exists\, p^{e_1} \overset{abs}{\sim} p^{e_2}, p^{e_2} \overset{abs}{\sim} p^{e_2} \tag{15}$$

Если условие выполняется только для одного события, но не выполняется для другого, то возможно существует иерархия типов событий. Однако данное исследование не затрагивает этот аспект анализа.

Таким образом, функция определения однотипных связей по признаку абсолютного сходства f^{sem} является отображением множества событий E и множество их свойств P во множество типов событий ET может быть выражена как:

$$f^{abs}(e^1, e^2) = \begin{cases} P^{e_1} \Delta P^{e_2} = \emptyset: \ et^{e_1} = et^{e_2} \\ \text{иначе: } et^{e_1} \neq et^{e_2} \end{cases}, \ f^{abs}: \{E, P\} \to ET, \quad (16)$$

где Δ - оператор симметрической разности множеств. Данное отображение является суръективным, поскольку для каждого элемента множества ET его полный прообраз не является пустым множеством:

$$\forall \, et \in ET: \{f^{abs^{-1}}(et)\} \neq \emptyset \quad (17)$$

Однотипный вид отношений между свойствами введен для типизации событий, что позволит в дальнейшем анализировать поведение отдельных информационных объектов по последовательностям типов событий каждого из них.

Следует отметить, что результат сравнения типов двух событий по абсолютному сходству является бинарным, то есть типы событий и их свойства либо эквивалентны, либо нет. Все остальные виды отношений задают характеристику связи между свойствами в количественном эквиваленте.

Предполагается, что свойства событий типизированы, а смысл каждого свойства однозначен. Однако смысл некоторых свойств может совпадать. *Однотипные отношения по признаку типовой эквивалентности* между свойствами событий задают множество типов свойств PT:

$$PT = \{pt_1, pt_2, \dots, pt_m\}, \quad (18)$$

где m – количество типов свойств событий. Функция определения подобных связей f^{type} опирается на следующую гипотезу: пересечение множеств значений V^{p_1} и V^{p_2} двух однотипных свойств p_1 и p_2, эквивалентных по типовой схожести, не является пустым множеством:

$$V^{p_1} \cap V^{p_2} \neq \emptyset \ \forall \, p_1 \overset{type}{\sim} p_2, \{p1, p2\} \subset P, \{V^{p_1}, V^{p_2}\} \subset V, \quad (19)$$

В идеальном случае, пересечение множеств значений V^{p_1} и V^{p_2}однотипных свойств p_1 и p_2, эквивалентных по типовой схожести, является каждым из множеств V^{p_1} и V^{p_2}. Функция определения данного вида отношений между свойствами задается выражением:

$$f^{type}(p1, p2) = \frac{|V^{p_1} \cap V^{p_2}|}{|V^{p_1} \cup V^{p_2}|} \quad (20)$$

Таким образом, данная функция является суръективным отображением множества свойств P и множеств их значений V в множество PT, поскольку множество всех прообразов для каждого типа свойств pt не пустое:

$$f^{type}: \{P, V\} \to PT, \forall\, pt \in PT : \left\{ f^{type^{-1}}(pt) \right\} \neq \emptyset \qquad (21)$$

В случае если свойство p не имеет эквивалентных свойств из P по типовому сходству, то множество его значений формирует отдельный тип свойств.

Опираясь на вышеописанную связь между свойствами и выражение 10 следует, что множество типов свойств PT событий безопасности E определяет множество типов характеристик XT объектов инфраструктуры I:

$$PT^L = XT^I, XT = \{xt_1, xt_2, \dots, xt_m\} \qquad (22)$$

то есть значения каждой характеристики x из X определены конкретным типом характеристик xt (типом свойств pt).

Разнотипные отношения между свойствами событий безопасности могут определяться различными признаками, такими как: взаимная используемость свойств в событиях и используемость по типам событий, взаимная уникальность значений свойств, коррелируемость значений свойств и многими другими. Однако описываемый далее подход структурного анализа событий затрагивает только *разнотипные отношения* между свойствами, определяющих их *эквивалентность по признаку взаимной используемости*.

Предполагается, что различные события безопасности E описывают различные действия, которые выполняют объекты-источники над объектами-целями из множества O и описываемые с помощью характеристик объектов X. Функция определения подобных разнотипных отношений f^{using} опирается на следующую гипотезу: исключительно-совместное использование свойств p_1 и p_2, задающих характеристики x_1 и x_2 одного или нескольких объектов из множества O, свидетельствует о принадлежности описываемых объектов к одному или нескольким типам объектов OT, определяемых их типами характеристик XT. Также предполагается, что абсолютный показатель используемости свойств свидетельствует об уровне описываемых объектов в иерархии типов объектов инфраструктуры.

Функция определения разнотипной связи по взаимному использованию f^{using} между свойствами событий p_1 и p_2 описывается следующим выражением:

$$f^{using}(p_1, p_2) = \frac{|E^{p_1} \cap E^{p_2}|}{|E^{p_1} \cup E^{p_2}|}, \{p_1, p_2\} \subset P, \{E^{p_1}, E^{p_2}\} \subset E \quad (23)$$

С учетом вышеописанного, функция f^{using} является отображением множеств событий *E* и их свойств *P* на множество типов объектов *OT*:

$$f^{using}: \{E, P\} \rightarrow OT \quad (24)$$

Стоит отметить, что сила связей между событиями, эквивалентность которых задается вещественным показателем, имеет диапазон значений от [0,1]. Следовательно, для установления факта наличия связи между свойствами (перевода к булевому результату), следует задавать некий порог *lim*, преодоление которого в большую сторону позволяет считать связь существенной. Данный параметр подробно исследовался для разнотипных отношений между свойствами по эквивалентности их использования и описан в разделе результатов.

В результате определенных элементов модели НЦИ КФС дальнейший этап процесса корреляции заключается в определении степени динамичности характеристик объектов для их связывания с онтологической моделью корреляции данных безопасности с условно-статичным содержимым. Очевидно, что данные процедуры основываются на взаимном сопоставлении элементов из двух моделей по наиболее статичным ресурсам (файлы в файловой системе, доменные имена, сетевые адреса и др.). Дальнейший анализ подразумевает мониторинг динамики состояния всех выделенных информационных объектов для формирования соответствующих моделей поведения и предсказания вероятности наступления конкретных событий (предупреждений, инцидентов) в заданный промежуток времени.

Предлагаемый подход корреляции также имеет ряд требований к исходным данным для корректного выполнения процедур адаптации:

6) Любой тип событий *et* не может содержать два абсолютно равнозначных свойства:

$$\forall \{p_1, p_2\} \in P^{ot} : p_1 \nsim p_2, \quad (25)$$

7) поскольку в данном случае невозможно однозначно отличить принадлежность характеристик *x* к описываемым объектам *O*.

8) Свойства событий однозначно характеризуют определенный тип информации, значения каждого свойства имеют единый формат представления, а семантика двух не однотипных свойств не может совпадать. Иначе, при использовании ненормализованных данных, применение подхода породит значительное количество ошибок второго рода при определении однотипных отношений между свойствами по типовой эквивалентности.

9) Полнота исходных данных должна быть достаточной для принятия решения о наличии той или иной связи между свойствами событий. То есть каждому типу события *et* должно соответствовать хотя бы одно из событий *E,* иначе тип события *et* не существует. Также каждое свойство *p* встречается хотя бы в одном из типов событий *ET*, а множество значений V^p каждого из свойств *P* не пустое. В противном случае, свойство *p* не существует.

Таким образом, ситуации, в которых указанные требования не соблюдены, выражают общее ограничение комплексной методики корреляции данных безопасности: выполнение методики подразумевает сбор, предварительную обработку, адаптацию к НЦИ и анализ только эмпирически полученных данных, то есть наблюдаемых в прошлом или настоящем.

Выводы по главе 2

1) Описана модель НЦИ, на основе которой предполагается обучение моделей поведения ее информационных объектов. Определение данной модели основано на анализе эмпирических данных с динамичным содержимым ЦИ КФС.

2) Разработана модель корреляции данных с условно-статичным содержимым на основе онтологического подхода. Анализ подобных данных (описаний уязвимостей, эксплойтов, продуктов и др.) подтвердили наличие взаимосвязей между информацией различной природы. Предлагаемая модель

позволяет извлекать знания с помощью заданных отношений между сущностями, тогда как экземпляры всех конечных сущностей в конечном счете отражают элементарные информационные статичные (условно) объекты каждого из источников подобной информации.

3) Разработана модель корреляции данных с условно статичным содержимым на основе ввода типизации для событий и их свойств, информационных объектов, их характеристик и взаимоотношений между ними.

4) Разработана комплексная методика корреляции данных безопасности с поддержкой автоматизированной адаптацией к НЦИ. Задача адаптации реализована за счет статистического и структурного анализа исходных данных с динамичным содержимым, а результатом ее выполнения является определение модели ЦИ КФС.

Глава 3 Реализация системы корреляции больших массивов гетерогенных данных безопасности и оценка ее эффективности

3.1 Реализация и оценка эффективности предлагаемого подхода корреляции данных безопасности

Описанный выше подход корреляции событий безопасности был практически реализован и экспериментально протестирован. Исходный код прототипа написан на языке Python 3.5 с использованием библиотек numpy, scipy и pandas, а результаты визуализированы с помощью GraphViz и модулей matplotlib, pyplot и seaborn. Для выполнения экспериментов использовалась вычислительная платформа с одним 6-ядерный процессор Intel(R) Xeon(R) CPU E5-2603 v3 @ 1.60GHz. и 64 GB RAM.

Пример связывания информации безопасности с условно-статичным содержимым по правилам разработанной онтологической модели приведен на рисунке 16. Рассматривается ситуация, когда за счет выполнения запроса осуществляется связывание разнотипной информации на основе логического вывода. Данный пример демонстрирует в действии ссылочный функционал связывания (связывания на основе перекрестных ссылок). Стоит отметить объектное свойство «имеетСвязьС», которое является верхнеуровневым в иерархии свойств-отношений между классами онтологии. Иными словами, все объектные свойства онтологической модели («реализуется», «реализует» и др.) так или иначе можно более обще представить данным свойством.

Пример демонстрирует связи между информацией безопасности по исходным данным. С помощью запросов к онтологической модели возможно получение ответом на следующие вопросы: (1) «Какая слабость реализуется посредством уязвимости, связанной с эксплойтом «*EDB-41443*»? »; (2) «Какой продукт содержит уязвимость, связанную с эксплойтом «*EDB-41443*»? ». Соответствующие запросы на языке DL Query выглядят следующим образом:

(1) «*Weakness and implementedBy some (Vulnerability and connectedWith some (Exploit and (localID value «EDB-41443»)))*»;

(2) «*Product and containsImplementationOf some (Vulnerability and connectedWith some (Exploit and (localID value «EDB-41443»)))*».

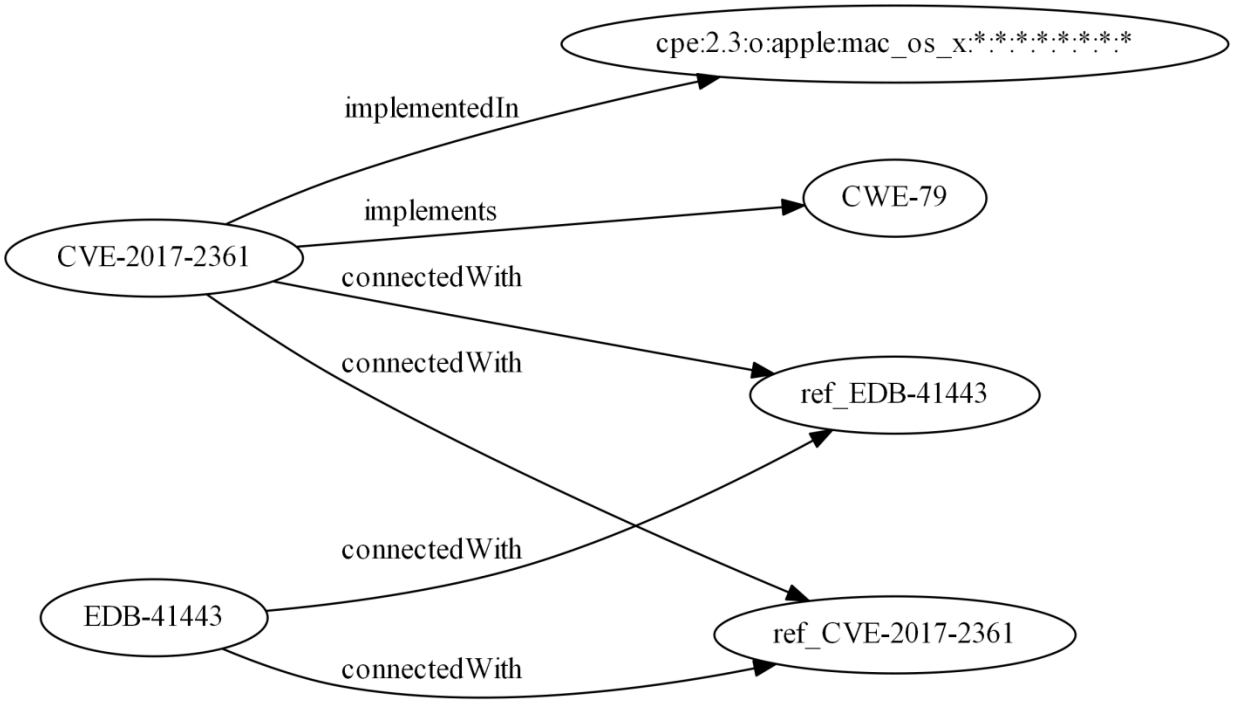


Рисунок 16 - Пример взаимосвязей информации безопасности с условно-статичным содержимым

Результатами выполнения запросов соответственно являются объекты: (1) слабости базы CWE «*CWE-79*» и (2) продукт «*cpe:2.3:o:apple:mac_os_x:*:*:*:*:*:*:*:*:**».

Описание онтологической модели на языке OWL (разновидность и профиль указан выше) с одной стороны достаточно семантически выразительно, а с другой ответы на вопросы будут получаться за полиномиальное время ($T(n)=O(n^k)$), в зависимости от количества используемых ресурсов и индивидов гибридного хранилища статичных данных.

Исходные данные безопасности с динамичным содержимым. В условиях неопределенной целевой инфраструктуры множества *E*, *P* и *V* изначально неизвестны. Пользуясь историческими записями событий, либо накапливая

информацию в реальном времени следует первоначально определить множества свойств *P* и их возможных значений *V*.

Исходными данными для проведения экспериментов являются события системного журнала безопасности хоста под управлением ОС Windows 8. Подсистема журналирования операционной системы была настроена на сбор максимального количества типов событий безопасности. Из данного набора данных изначально исключены свойства, определяющие временную привязку событий и их идентификаторы. Такие свойства являются достаточно уникальными (в пределах данного журнала) и определяют последовательность и порядок событий в журнале. Данные аспекты будут рассмотрены в будущих исследованиях. В итоге анализируемый журнал обладает следующими характеристиками:

1) Количество событий: ~ 6700000;
2) Количество заявленных типов событий: 44 (из более чем 250 заявленных [74]);
3) Количество свойств событий: 110;
4) Размер данных журнала: 7ГБ в формате XML, 1,25ГБ в формате CSV;
5) Время записи журнала: 36 дней.

Определение типов событий. В ходе анализа журнала было выявлено 37 уникальных наборов свойств, каждый из которых определяет отдельный тип событий. Дальнейший анализ показал, что 7 пар заявленных типов событий имеют равные множества используемых свойств. Соответствие данных типов представлено в таблице 6.

С одной стороны, полученные результаты свидетельствуют о различии типов событий по вариантам набора их свойств. С другой стороны, ряд типов событий, описывающих смежные действия, то есть действия одних и тех же типов объектов, имеют совпадающие множества свойств. В результате обнаружения однотипных отношений между свойствами событий по признаку абсолютной эквивалентности данные типы были объединены в один общий.

Таблица 6 - Соответствие типов событий с совпадающими наборами свойств

Тип события 1	Тип события 2
4670: Permissions on an object were changed	4907: Auditing settings on object were changed
4778: A session was reconnected to a Window Station	4779: A session was disconnected from a Window Station
4800: The workstation was locked	4801: The workstation was unlocked
4904: An attempt was made to register a security event source	4905: An attempt was made to unregister a security event source
4946: A change has been made to Windows Firewall exception list. A rule was added	4948: A change has been made to Windows Firewall exception list. A rule was deleted
5154: The Windows Filtering Platform has permitted an application or service to listen on a port for incoming connections	5158: The Windows Filtering Platform has permitted a bind to a local port
5156: The Windows Filtering Platform has allowed a connection	5157: The Windows Filtering Platform has blocked a connection

Таким образом, эмпирическая точность обнаружения типов событий за счет предлагаемого подхода корреляции событий безопасности составляет 84%. В дальнейших исследованиях следует также проанализировать возможность типизации событий по значениям их свойств.

Деление исходной выборки событий безопасности по типам имеет ряд практических преимуществ:

- верификация форматов поступающих на анализ событий. Позволяет регистрировать как новые типы событий, так и проверять на корректность уже зарегистрированные;
- возможность табличного представления данных для хранения. Упрощает и ускоряет доступ к информации за счет распределенного сбора и параллельной обработки с учетом большого объема исходных данных, минуя этап парсинга;
- предварительный отбор данных для обработки на основе выявленных связей между типами. В первую очередь анализу подлежат данные, являющиеся

описанием равнозначных и неравнозначных однотипных свойств между типами событий.

Определение типов свойств. В результате анализа исходных данных 38 свойств отнесены к 9 типам при пороге значимости связей *lim* равному 0,7. Следует определить нулевой тип, который обозначает нулевые свойства. Свойство является нулевым если принимает единственное значение, не характеризующий смысл свойства. Данный тип необходим в случае недостаточной нормализации данных, когда в определенных событиях значения некоторых свойств не может быть заполнено корректно. Таким образом, все свойства нулевого типа во всех обработанных событиях принимают нулевое значение. В данном случае таким значением является символ «-«, а количественный показатель типовой эквивалентности для каждой пары из совокупности нулевых свойств равен 1. В описанных исходных данных было выделено 6 нулевых свойств: «TransmittedServices», «CommandLine», «FileName», «TaskContentNew», «LinkName», «Conditions».

К каждому из остальных типов свойств (кроме нулевого) отнесено по 2 свойства, эквивалентных по типу. Выделенные свойства и их интерпретированные типы (по возможности) представлены в таблице 7. Один из типов не удалось интерпретировать, поскольку неясна связь между его свойствами.

Таблица 7 - Свойства, связанные по типовой эквивалентности и их типы

Interpreted properey type	Properties
‘RemoteID’	‘RemoteMachineID’, ‘RemoteUserID’
‘Workstation’	‘WorkstationName’, ‘Workstation’
???	‘AccountDomain’, ‘ThreadID’
‘HandleId’	‘SourceHandleId’, ‘HandleId’
‘ProcessName’	‘ProcessName’, ‘NewProcessName’
‘LogonGuid’	‘LogonGuid’, ‘TargetLogonGuid’
‘ProcessId’	‘NewProcessId’, ‘ProcessId’
‘UserSid’	‘TargetUserSid’, ‘SubjectUserSid’
‘Target’	‘TargetServerName’, ‘TargetInfo’

Стоит отметить, что при значении порога *lim* = 1, подход позволяет выделить только 5 типов (включая нулевой), которые задают типизацию для 14 свойств. В свою очередь, при *lim* = 0.5 выделяется 14 типов, причем к некоторым из них относится более двух свойств. Однако, с уменьшением порога значимости связей по типовой эквивалентности возрастает количество коллизий, когда одно свойство относится сразу к двум типам событий, что противоречит второму требованию к исходным данным.

На рисунках 17-19 представлены наиболее значимые типы свойств, сформированные за счет связей по эквивалентности типов без задания порога *lim*. Узлами графов обозначены свойства типов событий, а дугами – взвешенные связи между свойствами с указанием количественного показателя отношения. Связи, показатель типовой эквивалентности которых приближен к 0 – не отображаются. Далее, каждый из примеров сформированных типов свойств событий будет рассмотрен отдельно.

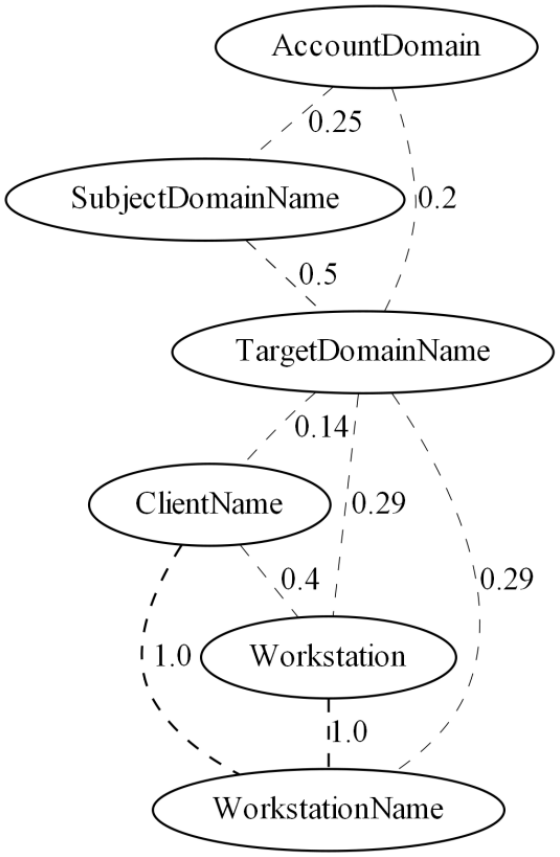


Рисунок 17 - Свойства типа «домен»

Рисунок 17 описывает свойства типа «домен» и количественное значение их отношения по типовой эквивалентности. Связи между свойствами отчетливо делятся на сильные и слабые. Очевидно, что при задании порога значимости *lim* = 0.3 для данных связей, представленный тип поделился бы еще на 3 типа свойств.

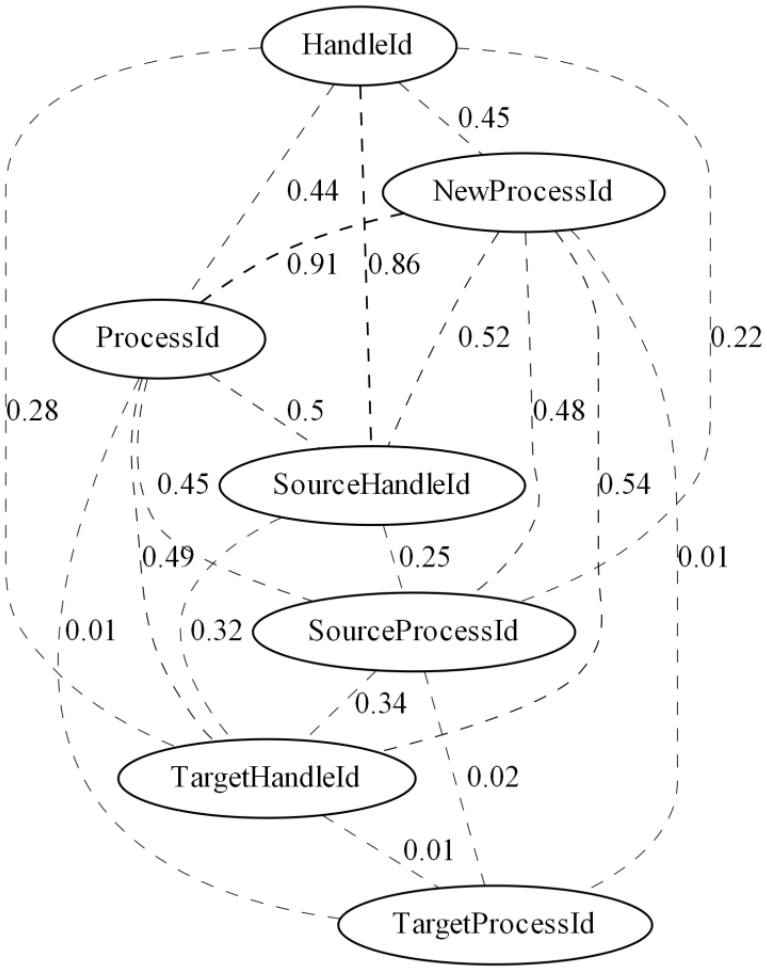


Рисунок 18 - Свойства типа «идентификатор процесса-хэндла»

Рисунок 18 описывает тип свойств «идентификатор процесса-хэндла», который является показательным примером объединения двух типов свойств, а также ошибки первого рода. Во-первых, очевидно, что представленный тип содержит 2 отдельных типа: «идентификатор процесса» и «идентификатор хэндла». Подобная ошибка связана с тем, оба типа свойства выражены количественными значениями ограниченного диапазона, что нарушает второе требование к исходным данным. Во-вторых, слабая связь свойства *TargetProcessId* с другими свойствами типа *ProcessId*, что является ошибкой

первого рода, объясняется низкой вариативностью свойства *TargetProcessId*. Другими словами, даже если все значения свойства *TargetProcessId* являются , например, подмножеством значений свойства *ProcessId*, при большом преобладании мощности второго над первым значение связи будет стремиться к 0.

В дополнении, было обнаружено следующее нарушение требования к исходным данным 2: значения свойства ProcessId выражены целочисленным значением как в десятичной системе счисления, так и в шестнадцатеричной с префиксом «0x», что свидетельствует о несогласованном ведении журнала безопасности ОС Windows.

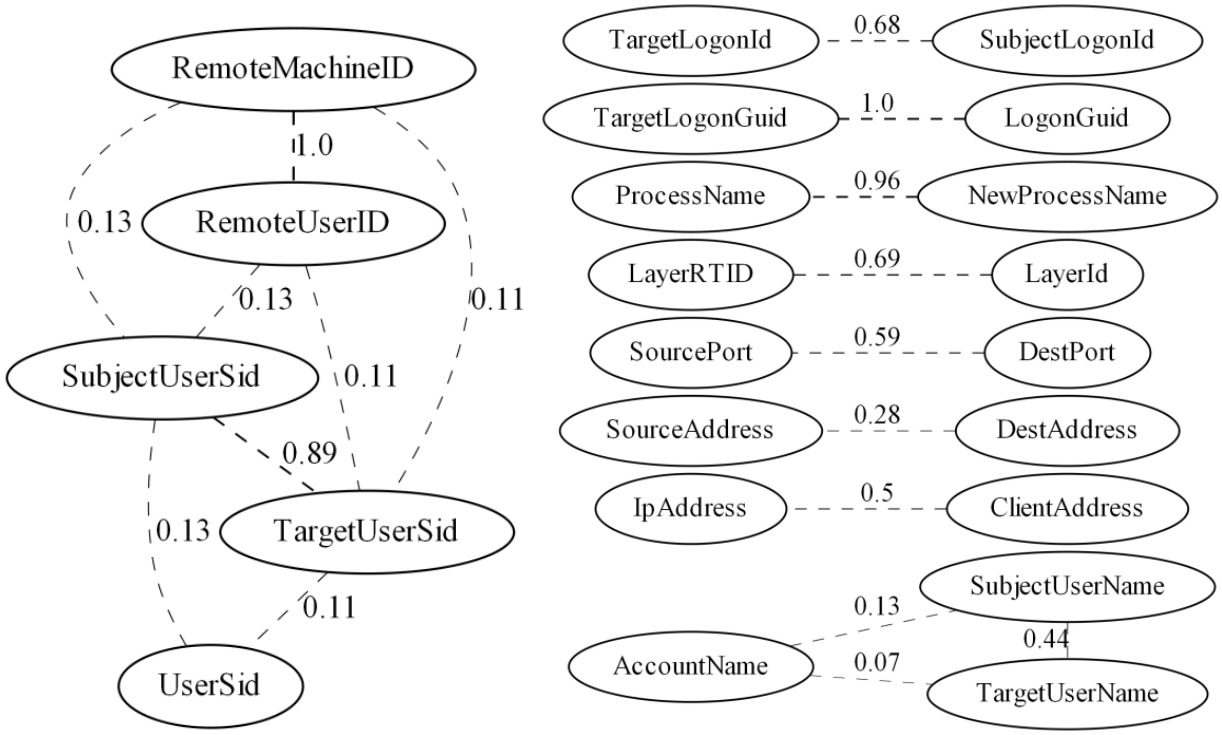


Рисунок 19 - Свойства типа «идентификатор пользователя и другие типы свойств»

На рисунке 19 (слева) представлена группа свойств, определяющих тип «идентификатор пользователя», а справа – прочих связанных по типовой эквивалентности свойств событий безопасности. Связи между свойствами типа «идентификатор пользователя» также как и в первом примере заметно отличаются по силе. Однако в данной ситуации представленные свойства действительно являются однотипными, а само явление скорее является множественным случаем ошибки первого рода, причины которой описаны в примере 2.

Таким образом, экспериментально подтверждена возможность применения отношений между свойствами по типовой эквивалентности для задачи выявления типов свойств событий безопасности в процессе корреляции. 57 свойств (40% от общего количества, не считая свойств нулевого типа) были идентифицированы как однотипные по отношению к одному и более неравнозначным свойствам. При дальнейшей оценке, 7 свойств (12% от идентифицированных) были выделены явно ошибочно. В свою очередь, правильно выделенные однотипные свойства образуют 46 косвенных отношений (32% по сравнению с равнозначными свойствами) без учета их транзитивности.

На представленных рисунках 16-18 не все показатели схожести свойств имеют достаточное значение (lim>0.1) для автоматизированного принятия решения об их однотипности. Однако такие отношения все же были приняты как корректные на основе экспертной оценки их очевидного сходства.

В результате проведенного анализа были установлены существенные замечания к выполнению задачи поиска однотипных свойств эквивалентных по типу данных, а именно:

1) Разнообразие исходных данных пропорционально качеству и количеству обнаруженных косвенных связей. Увеличение разнообразия может быть достигнуто путем анализа большего объема исходных данных, а также сбора информации от большего количества источников. Данное условие необходимо при адаптации предлагаемого подхода к ЦИ.

2) Использование количественного типа данных для описания свойств событий приводит к увеличению некорректно идентифицированных однотипных неравнозначных свойств (за исключением случаев, когда семантика свойства определяется множеством вещественных чисел). Например, в журнале безопасности ОС Windows подобным образом описываются такие свойства, как длина ключа («KeyLenght»), номера сетевых портов («IpPort», «DestPort», «SourcePort»), тип протокола («Protocol») и многие другие. В тоже время, использование количественного типа для описания значений таких типовых свойств, как тип протокола («Protocol»), тип изменения («ChangeType») и других,

искажает результат определения косвенной однотипной связи. Такое искажение также наблюдается у свойств со ссылочным типом данных. Для преодоления возможной неопределенности, типовые свойства необходимо представлять в строковом виде (например, «Protocol=udp»), а ссылочные значения должны быть разыменованными. При невозможности отказа от численного представления значений некоторых свойств, следует прибегать к дополнительному анализу их значений. Данный анализ выполняется за счет вычисления таких характеристик, как диапазон возможных значений, кучность значений и др.

3) Несогласованное присвоение идентификаторов описываемым в событиях объектам также приводит к увеличению числа ложных косвенных связей. Таким образом, однотипные свойства (например, различные идентификаторы процесса) должны иметь уникальный формат идентификатора среди всех множеств неравнозначных свойств.

4) Несогласованность употребления нулевых значений негативно влияет на поиск существующих однотипных неравнозначных свойств, а также может приводить к некорректному определению отношений между ними. Например, в обработанном наборе данных нулевые значения представлены в виде: «-», «null», «0», «0x0», «{00000000-0000-0000-0000-000000000000}» и др. В ходе анализа, большинство подобных значений были искусственно исключены при вычислении показателя схожести. Однако, данный факт усложняет выполнение автоматизированной адаптации.

5) Использование различных типов данных для множества значений одного свойства является грубой ошибкой и негативно влияет на точность процесса корреляции при учете смысла значений свойств событий. Например, в журнале безопасности ОС Windows свойство SourceAddress может принимать значения как IP-адреса (например, «127.0.0.1»), так и MAC-адреса (например, «1234:5678:abcd»). Подобное использование увеличивает время анализа журнала экспертом и значительно усложняет автоматизированный анализ событий. В частности, пересечение разных типов данных в множестве значений одного свойства уменьшает его показатель схожести по отношению к другим свойствам.

6) Увеличение точности определения косвенных однотипных связей возможно за счет использования шаблонов распознавания типов значений, например, регулярных выражений. В данном случае можно определять такие типы данных, как IP-адрес, MAC-адрес, путь в файловой системе, URL-ссылка, GUID-идентификатор и другие.

К сожалению, большинство из перечисленных замечаний могут быть устранены только разработчиками подсистем журналирования. Однако так как большинство выявленных проблем касаются этапа нормализации данных, также возможно решение на основе применения соответствующих правил обработки исходной информации.

Определение типов объектов. Как было отмечено ранее, определение типов объектов по свойствам, эквивалентных по взаимной используемости, основывается на гипотезе, что одинаковая используемость свойств событий свидетельствует об описании характеристик одного или нескольких типов объектов. На рисунке 20 представлена гистограмма использования свойств относительно общего количества событий. Отображены только те свойства, используемость которых в событиях журнала превышает 3%, то есть если свойство встречается менее чем в 3% событий, то на данном графике его нет. Отчетливо видно, что отдельные группы свойств имеют равный, либо очень близкий по значению, показатель используемости. Таким образом, выдвигаемая гипотеза предварительно подтверждается.

В результате эксперимента по определению разнотипных связей между свойствами на основе показателя их совместного использования было выделено 18 групп свойств, а их общее количество равно 60. Стоит отметить, что значение предела значимости связи в данном случае было 1. Другими словами свойства из одной группы используются в одинаковом количестве типов событий.

Наиболее значимые и интерпретируемые типы объектов вместе со свойствами, которые их определяют, представлены на рисунке 21. На данном рисунке в каждой группе событий также отмечено среднее значение общего использования составляющих группу свойств, а сами группы упорядочены по

данному показателю. Однако, с показателем значимости связи *lim* = 1, указанный средний уровень используемости для группы будет соответствовать используемости каждого из свойств.

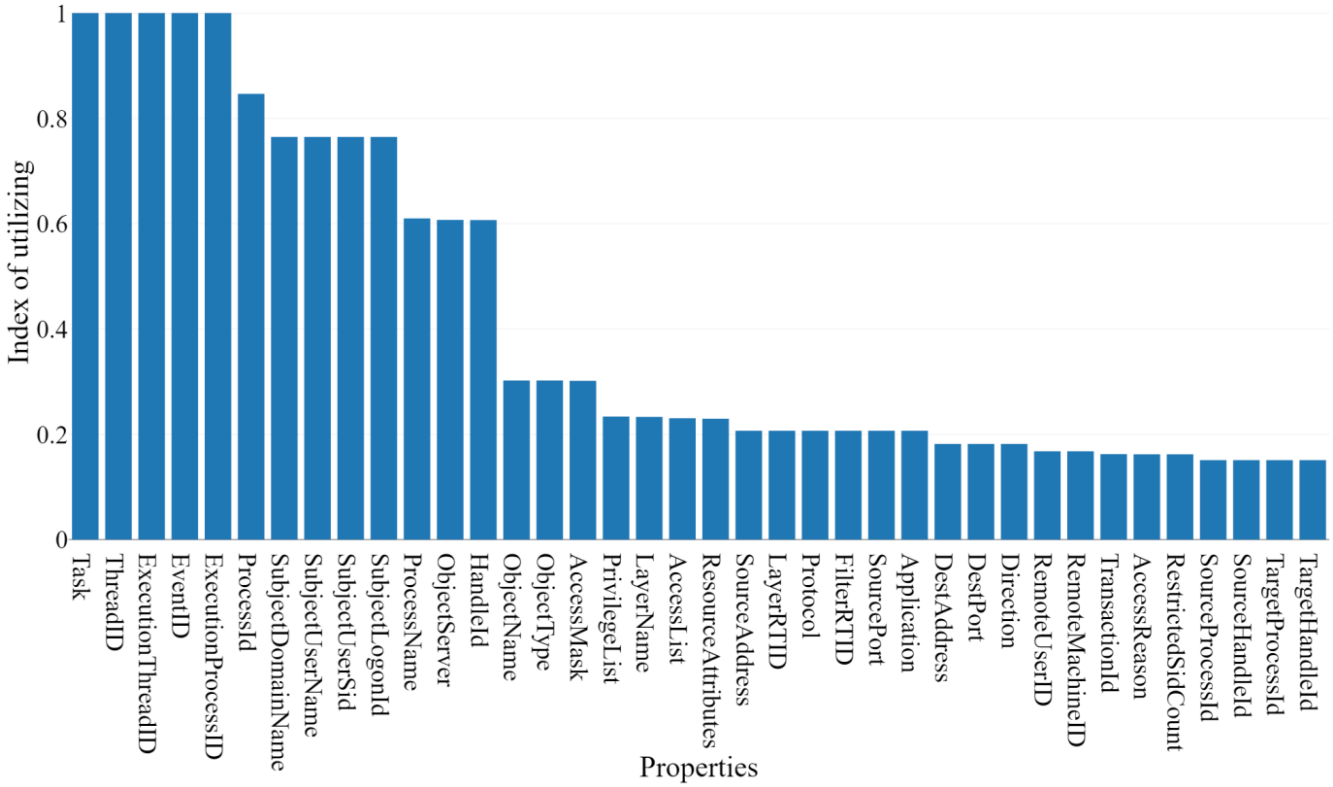


Рисунок 20 - Гистограмма используемости свойств событий в журнале безопасности OS Windows

Отдельного внимания заслуживают объекты нулевого типа, которые встречаются во всех типах событиях. Строго говоря, данные свойства задают нулевой тип событий, поскольку даже пустое событие как структура данных имеет определенный заголовок со служебной информацией, например, тип события «EventID» или задача «Task». Стоит отметить, что факт наличия нулевого типа объекта (события) в анализируемой инфраструктуре можно установить за счет проверки показателя используемости, поскольку для подобных свойств он должен быть равен 1.

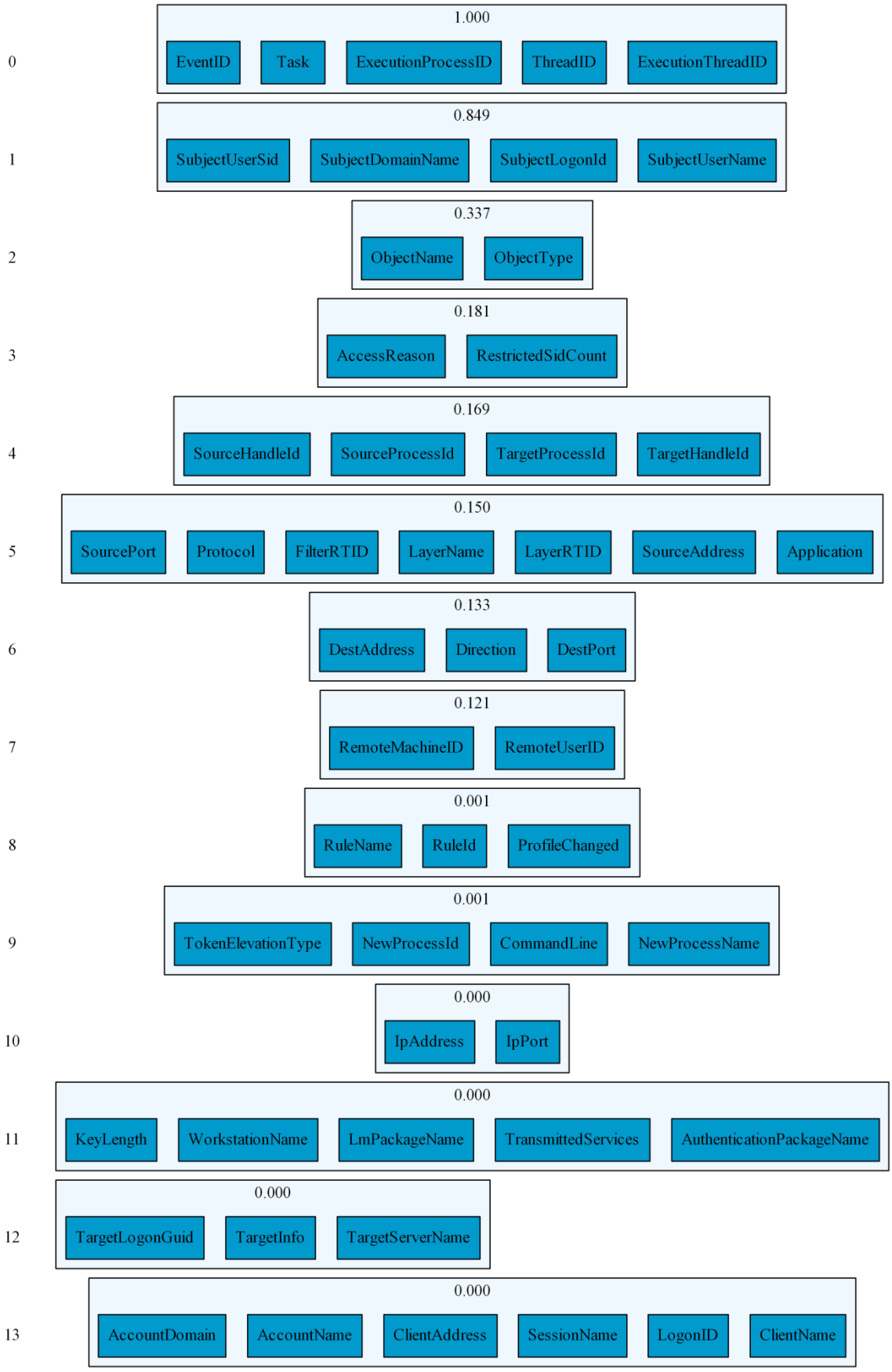


Рисунок 21 - Типы объектов ОС Windows.

Рассмотрев сформированные типы объектов следует заключить, что большинство их них (за исключением нулевого уровня) являются достаточно семантически выраженными. Очевидно, что представленные группы свойств действительно формируют типы высокоуровневых объектов. Так, группы свойств 1 и 2 формируют типы объектов «Субъект» и «Объект» соответственно. В то же время, группа 4 явно должна быть поделена на 2 типа объекта «Источник» и «Цель». Вероятно более глубокая типизация объектов, например, на основе их поведения, позволит избежать подобных неточностей. Данное направление будет рассматриваться в дальнейших исследованиях.

Полученные типы объектов описываются 54% всех имеющихся свойств событий в журнале безопасности с порогом значимости связи между свойствами $lim = 1$. Однако, с таким значением порога был пропущен ряд достаточно используемых свойств. Данный случай связан с некоторыми погрешностями, возникающими, например, при нормализации исходных данных, когда количество событий с определенным нормализованным свойством незначительно меняется. Такое изменение уже не гарантирует идентичного взаимного использования свойств, следовательно, такие случаи должны обрабатываться с помощью порога значимости связей между свойствами.

В ходе анализа показателя порога значимости было установлено, что даже незначительное допущение совместной используемости свойств событий приводит к увеличению количества типов объектов. При использовании порога значимости $lim = 0{,}99$ для определения отношений между свойствами по типовой эквивалентности, был идентифицирован еще один тип объектов, который определен тремя свойствами: «ProcessName», «HandleId» и «ObjectServer». Также верны следующие утверждения: изменение порога значимости приводит к изменению количества и состава свойств типов объектов, а также к появлению коллизий. В данном случае коллизия означает отнесение свойства сразу к нескольким типам объектов.

Общее время выполнения задач описанного подхода корреляции составляет не больше двух минут. Самыми продолжительными операциями являются:

(1) загрузка данных в оперативную память для последующей обработки, поскольку в данном случае скорость загрузки ограничивается скоростью считывания данных с жесткого диска; (2) определение отношений эквивалентности по взаимной используемости. Среднее время выполнения обеих операций составляет около 30 секунд (вторая операция выполнялась параллельно в 6 потоках). Учитывая, что исходные данные являются записями событий достаточно длительного отрезка времени, время выполнения анализа можно считать приемлемым. В целом, предлагаемый подход корреляции событий на основе их структурного анализа показал достаточно высокие результаты, а большинство случаев, в которых он проявил себя нестабильно, вероятнее всего вызвано некоторыми отклонениями исходных данных от предъявляемых требований.

3.2 Предложения по развитию и использованию системы корреляции для задач обеспечения безопасности КФС

Разработанную систему корреляции данных безопасности в дальнейшем предполагается использовать для проактивного мониторинга и получения предварительного состояния безопасности защищаемой КФС за счет применения интеллектуальных методов. В данном случае в качестве учителя будут выступать всевозможные средства защиты информации, выявляющие инциденты безопасности, которые в свою очередь являются высокоуровневыми событиями. Таким образом, возможные критичные состояния инфраструктуры будут определяться за счет предшествующих инцидентам низкоуровневых событий. Очевидно, что для применения подобного подхода необходим этап обучения, в рамках которого анализируется нормальное поведение НЦИ КФС, а также определяются типы событий, вероятность возникновения которых следует вычислять в реальном времени. Также следует учесть, что процесс обучения для проактивной корреляции информации в киберфизических инфраструктурах по

полученной модели должно достигаться за счет моделирования критичных ситуаций (инцидентов) физического уровня.

Дальнейшими задачами по развитию онтологической модели гибридного хранилища информации безопасности и предлагаемого подхода корреляции данных безопасности с условно-статичным содержимым в целом являются:

1) добавления сущностей описания конфигурации целевой инфраструктуры и их объектных свойств для установления связей между различными типами информации безопасности;
2) уточнение модели для концептов слабостей и шаблонов атак, позволяющее более гибко использовать иерархические структуры их классификаций;
3) разработка функционала определения прямого и обратного соответствия продуктов и их обобщающих записей (по признакам версий, модификаций, редакций и т.д.).

Выводы по главе 3

1) Описан пример реализации процедуры корреляции между информацией с условно-статичным содержимым за счет разработанной онтологической модели и языка запросов дескриптивной логики.
2) Описаны результаты выполнения процедур адаптации системы корреляции к НЦИ КФС по экспериментальными эмпирическими данными с динамичным содержимым.
3) Оценена сложность выполнения разработанных процедур корреляции данных безопасности.

Заключение

В результате проделанной работы были исследованы этапы нормализации, предварительной обработки и анализа данных безопасности в процессе корреляции. Предложен подход, ориентированный как на адаптацию к целевой инфраструктуре в условиях неопределенности КФС, так и на связывание ее статических и динамически объектов разной природы.

На основе проведенного исследования был разработан модельно-методический аппарат для реализации адаптивного процесса корреляции данных безопасности, суть которого заключается в определении модели НЦИ КФС и вычисления состояния защищенности ее активов. Основное преимущество предлагаемого подхода заключается в преобразовании категорийной информации в численные и нормированные показатели, а также в вычислении стато-динамических показателей характеристик объектов.

Дальнейшая работа подразумевает накопление информации из разнородных источников и ее интеграцию в виде графовой модели связей для выполнения интеллектуального анализа данных. Последующее выполнение поведенческого анализа с помощью данной модели будет основано на определение наиболее важных типов информационных объектов и зависимостей их поведения от случайных факторов, таких как поведение пользователя и внешние воздействия. Реализация предлагаемых механизмов анализа данных учитывает требование к автоматической или автоматизированной адаптации к НЦИ КФС. Практические результаты подтверждают справедливость предлагаемого подхода для его непосредственного применения для задачи корреляции данных безопасности к КФС.

Перечень используемых сокращений и обозначений

АСУ ТП - Автоматизированная Система Управления Технологическим Процессом

ИС - интеллектуальный сенсор

КВО – критически важный объект

КФС – кибер-физическая система

КФИ – кибер-физическая инфраструктура

НЦИ – неопределенная целевая инфраструктура

ОС – операционная система

ПАО – программно-аппаратное обеспечение

ПЛК - программируемый логический контроллер

ПО – программное обеспечение

РФ – Российская Федерация

ТУД - терминал удаленного доступа и управления

ЦИ – целевая инфраструктура

ЧМИ - человеко-машинный интерфейс

APT – передовая постоянная угроза (Advanced Persistent Threat)

CAPEC – «Общее перечисление и классификация шаблонов атак» (Common Attack Pattern Enumeration and Classification)

CCE – «Общее перечисление конфигураций» (Common Configuration Enumeration)

CPE – «Общее перечисление платформ» (Common Platform Enumeration)

CRE – «Общее перечисление защитных мер» (Common Remediation Enumeration)

CVE – «Общие уязвимости и дефекты» (Common Vulnerabilities and Exposures)

CVRF – «Фрэймворк учета общих уязвимостей» (Common Vulnerability Reporting Framework)

CVSS – «Общая система оценки уязвимостей» (Common Vulnerability Scoring System)

CWE – «Общее перечисление слабых мест» (Common Weakness Enumeration)

DL – дескриптивная логика (Descriptive Logic)

FIRST – Форум групп безопасности и реагирования на инциденты (Forum of Incident Response and Security Teams)

IDS – Intrusion Detection System

IoT – Интернет вещей (Internet of Things)

IPS – Intrusion Prevention System

NVD – «Национальная база уязвимостей» (National Vulnerability Database)

OSVDB – «База данных уязвимостей с открытым котом» (Open Source Vulnerability DataBase)

OVAL – «Открытый язык спецификации уязвимостей и оценки» (Open Vulnerability and Assessment Language)

OWL – язык веб-онтологий (Ontology Web Language)

RL – язык правил (Rule Languages)

SCADA – диспетчерское управление и сбор данных (Supervisory Control And Data Acquisition)

SCAP – протокол автоматизации управления данными безопасности (Security Content Automation Protocol)

SIEM – системы мониторинга безопасности и управления инцидентами (Security Information and Events Management)

US-CERT – компьютерная группа реагирования на чрезвычайные ситуации (United states computer emergency readiness team)

XML – расширяемый язык разметки (Extensible Markup Language)

Список литературы и электронных ресурсов

1) Kotenko I.V., Chechulin A.A. A Cyber Attack Modeling and Impact Assessment Framework // Proceedings of 5th International Conference on Cyber Conflict 2013 (CyCon 2013). 2013. pp. 119–142.
2) Kotenko I.V., Polubelova O.V., Saenko I.V. The Ontological Approach for SIEM Data Repository Implementation // 2012 IEEE International Conference on Green Computing and Communications, Conference on Internet of Things, and Conference on Cyber, Physical and Social Computing. Besançon, France, November 20-23, 2012. Los Alamitos, California. IEEE Computer Society. 2012. pp. 761–766.
3) Федорченко А.В., Чечулин А.А., Котенко И.В. Исследование открытых баз уязвимостей и оценка возможности их применения в системах анализа защищенности компьютерных сетей // Информационно-управляющие системы, 2014, №5, С.72-79.
4) Федорченко А.В., Чечулин А.А., Котенко И.В. Аналитический обзор открытых баз уязвимостей программно-аппаратного обеспечения // Проблемы информационной безопасности. Компьютерные системы, № 3, 2014. С.131-135.
5) Федорченко А.В., Чечулин А.А., Котенко И.В. Построение интегрированной базы уязвимостей // Известия высших учебных заведений. Приборостроение, Т.57, № 11, 2014. ISSN 0021-3454. С.62-67.
6) Федорченко А.В., Котенко И.В., Чечулин А.А. Разработка сервиса доступа и управления интегрированной базой уязвимостей // Безопасность информационных технологий, № 4, 2015. С.26-32.
7) Федорченко А.В., Левшун Д.С., Чечулин А.А., Котенко И.В. Анализ методов корреляции событий безопасности в SIEM-системах. Часть 1 // Труды СПИИРАН. 2016. Вып. 4(47). С.5-27.

8) Федорченко А.В., Левшун Д.С., Чечулин А.А., Котенко И.В. Анализ методов корреляции событий безопасности в SIEM-системах. Часть 2 // Труды СПИИРАН. 2016. Вып. 6(49). С.208-225.

9) 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).

10) Федорченко А.В., Котенко И.В. Корреляция информации в SIEM-системах на основе графа связей типов событий. Информационно-управляющие системы, №1, 2018. С.58-67.

11) Евгений Дружинин, Илья Карпов, Евгений Гнедин, Иван Бойко, Юлия Симонова. Безопасность АСУ ТП в цифрах. Positive Technologies. 2016. 20 с. [Электронный ресурс]. – Режим доступа: https://www.ptsecurity.com/upload/ptru/analytics/ICS-Vulnerability-2016-rus.pdf, 20.01.17.

12) Stuxnet. [Электронный ресурс]. – Режим доступа: http://www.securitylab.ru/news/tags/Stuxnet/, 20.01.17.

13) Анализ кода Stuxnet. Научный центр «НАУЦИЛУС». Москва. 2011. [Электронный ресурс]. – Режим доступа: http://www.phocus-scada.com/rus/pub/Stuxnet-CodeAnalys-rus.pdf, 20.01.17.

14) Анатолий Ализар. Подробности о беспрецедентном взломе электрической сети Украины. [Электронный ресурс]. – Режим доступа: https://geektimes.ru/post/272232/, 20.01.17.

15) Вирус, живущий исключительно в ПЛК. [Электронный ресурс]. – Режим доступа: https://habrahabr.ru/post/302276/, 20.01.17.

16) Взлом химического завода. [Электронный ресурс]. – Режим доступа: https://blog.kaspersky.ru/hacking-chemical-plant/8948/, 20.01.17.

17) Охтилев М.Ю, Соколов Б.В., Юсупов Р.М. Интеллектуальные технологии мониторинга и управления структурной динамикой сложных технических объектов. М.: Наука, 2006. 410 С.

18) Растригин Л.А. Современные принципы управления сложными объектами. М.: Сов. радио, 1980. 232 С.

19) С. Бир. Кибернетика и менеджмент. М: КомКнига, 2006. 280 С.

20) Andrey Fedorchenko, Igor Kotenko, and Didier El Baz. Correlation of security events based on the analysis of structures of event types // The 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS'2017). 21-23 September, 2017, Bucharest, Romania. P.270-276.

21) Kruegel C., Valeur F., Vigna G. Intrusion Detection and Correlation: Challenges and Solutions // University of California, Santa Barbara, USA: Springer. 2005.

22) Федорченко А.В., Дойникова Е.В., Чечулин А.А. Анализ источников данных и их форматов для систем аналитической обработки информации и событий безопасности // I Международная научно-техническая и научно-методическая конференция "Актуальные проблемы инфотелекоммуникаций в науке и образовании" (АПИНО-2017). 01-02 марта 2017 г. Сборник научных статей. Том 2. СПб., 2017.

23) Andrey Fedorchenko, Igor Kotenko, Elena Doynikova, Andrey Chechulin. The ontological approach application for construction of the hybrid security repository // XX International Conference on Soft Computing and Measurements (SCM'2017). IEEE Xplore, 2017. P.525-528.

24) Igor Kotenko, Andrey Chechulin, Elena Doynikova, Andrey Fedorchenko. Ontological hybrid storage for security data. Proceedings of the 11th International Symposium on Intelligent Distributed Computing - IDC'2017, Belgrade, Serbia, 11–13 October 2017. Springer-Verlag, Studies in Computational Intelligence, 2017, P. 159-171.

25) Котенко И.В., Дойникова Е.В. Анализ протокола автоматизации управления данными безопасности SCAP // Защита информации. Инсайд, 2012, № 2, С.56-63.

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

27) MITRE [Электронный ресурс] // URL: https://www.mitre.org/capabilities/cybersecurity/overview/cybersecurity-resources/standards (дата обращения 27.03.2017).

28) Nation Vulnerability Database https://nvd.nist.gov/ (дата обращения 27.03.2017).

29) Common Vulnerability Scoring System [Электронный ресурс] // URL: https://www.first.org/cvss (дата обращения 27.03.2017).

30) Open Source Vulnerability Data Base [Электронный ресурс] // URL: https://blog.osvdb.org/ (дата обращения 27.03.2017).

31) IBM X-Force Threat Intelligence [Электронный ресурс] // URL: https://www-03.ibm.com/security/xforce/ (дата обращения 27.03.2017).

32) Security Focus [Электронный ресурс] // URL: http://www.securityfocus.com/ (дата обращения 27.03.2017).

33) Common Vulnerability Reporting Framework [Электронный ресурс] // URL: http://www.icasi.org/cvrf/ (дата обращения 27.03.2017).

34) Liu G., Mok A.K., Yang E.J. Composite Events for Network Event Correlation // IEEE/IFIP International Symposium on Integrated Network Management. 1999. pp. 247–260.

35) Котенко И.В., Саенко И.Б., Полубелова О.В., Чечулин А.А. Технологии управления информацией и событиями безопасности для защиты компьютерных сетей // Проблемы информационной безопасности. Компьютерные системы. 2012. №2. С. 57–68.

36) Котенко И.В., Саенко И.Б., Полубелова О.В., Чечулин А.А. Применение технологии управления информацией и событиями безопасности для защиты информации в критически важных инфраструктурах // Труды СПИИРАН. 2012. Вып. 1 (20). С. 27–56.

37) Котенко И.В., Саенко И.Б., Чечулин А.А. Проактивное управление информацией и событиями безопасности в информационно-телекоммуникационных системах // Вопросы радиоэлектроники. 2014. Том 3. №1. С. 170–180.

38) Jakobson G., Weissman M.D. Alarm correlation // IEEE Network. 1993. no. 7(6). pp. 52 59.

39) Guerer D.W., Khan I., Ogler R., Keffer R. An artificial intelligence approach to network fault management. SRI International. 1996. p 10.

40) Tiffany M. A survey of event correlation techniques and related topics. URL: http://www.tiffman.com/netman/netman.html (дата обращения: 26.04.2016).

41) Hasan M. A conceptual framework for network management event correlation and filtering systems // Proceedings of the Sixth IFIP/IEEE International Symposium on Integrated Network Management. 1999. pp. 233–246.

42) Zurutuza U., Uribeetxeberria R. Intrusion Detection Alarm Correlation: A Survey // Proceedings of IADAT International Conference on Telecommunications and computer Networks. 2004. pp. 1–3.

43) Sadoddin R., Ghorbani A. Alert Correlation Survey: Framework and Techniques // Proceedings of 2006 International Conference on Privacy, Security and Trust: Bridge the Gap Between PST Technologies and Business Services (PST`06). 2006. Article no. 37.

44) Hanemann A., Marcu P. Algorithm Design and Application of Service-Oriented Event Correlation // Proceedings of Conference BDIM 2008, 3rd IEEE/IFIP International Workshop on Business-Driven IT Management. 2008. pp. 61–70.

45) Muller A. Event Correlation Engine. Master`s Thesis. Swiss Federal Institute of Technology Zurich. 2009. 165 p.

46) Limmer T., Dressler F. Survey of event correlation techniques for attack detection in early warning systems. Tech report. University of Erlangen, Dept. of Computer Science 7. 2008. 37 p.

47) Dadkhah S., Shoja M.R.K., Taheri H. Alert Correlation through a Multi Components Architecture // International Journal of Electrical and Computer Engineering (IJECE). 2013. vol. 3. no. 4. pp. 461–466.

48) Elshoush H.T., Osman I.M. Alert correlation in collaborative intelligent intrusion detection systems — A survey // Applied Soft Computing. 2011. pp. 4349–4365.

49) Ning P., Xu D. Correlation analysis of intrusion alerts // Intrusion Detection Systems: series Advances in Information Security. Springer, 2008. vol. 38. pp. 65–92.

50) Ahmadinejad S.H., Jalili S., Abadi M. A hybrid model for correlating alerts of known and unknown attack scenarios and updating attack graphs // Computer Networks. 2011. no. 55. pp. 2221–2240.

51) Elshoushand H.T., Osman I.M. An improved framework for intrusion alert correlation // Proceedings of World Congress on Engineering 2012 (WCE 2012). 2012. vol. 1. pp. 518–524.

52) Andrey Fedorchenko, Igor Kotenko and Andrey Chechulin. Integrated repository of security information for network security evaluation. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), Vol.6, No.2, June, 2015. P.41-57.

53) Horridge, M. A Practical Guide To Building OWL Ontologies Using Protege 4 and CO-ODE Tools, The University Of Manchester, 2011.

54) Protege wiki website. Protege user documentation. https://protegewiki.stanford.edu/wiki/Main_Page. Accessed on January 14, 2018.

55) W3C website. Web ontology language overview. https://www.w3.org/TR/owl-features. Accessed on January 14, 2018.

56) Vulners [Электронный ресурс] // URL: https://vulners.com/ (дата обращения 27.03.2017).

57) Andrey Fedorchenko, Igor Kotenko and Andrey Chechulin. Design of integrated vulnerabilities database for computer networks security analysis. 23th Euromicro International Conference on Parallel, Distributed, and Network-Based

Processing (PDP 2015). Turku, Finland, March, 2015. IEEE Computer Society. 2015. P.559-566.

58) Elahi G., Yu E., Zannone N. A modeling ontology for integrating vulnerabilities into security requirements conceptual foundations // Proc. 28th International Conference on Conceptual Modeling (ER-2009), Springer-Verlag Berlin, Heidelberg, 2009, pp. 99-114.

59) Guo M., Wang J. An ontology-based approach to model common vulnerabilities and exposures in information security // Proceedings of the 2009 ASEE SE Section Conference, 2009, 10 p.

60) Guo M., Wang J. Security data mining in an Ontology for vulnerability management // Proceedings of the 2009 International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing, 2009, pp. 597-603.

61) Kotenko I., Saenko I., Polubelova O., Chechulin A. Design and implementation of a hybrid ontological-relational data repository for SIEM systems // Future internet, vol.5, no.3, 2013, pp. 355-375.

62) Kotenko I., Saenko I., Polubelova O. and Doynikova E. The ontology of metrics for security evaluation and decision support in SIEM systems // Proc. of the 8th International Conference on Availability, Reliability and Security (ARES 2013), September 2013. Regensburg, Germany. IEEE Computer Society, 2013, pp. 638-645.

63) Kotenko I., Chechulin A. Attack Modeling and Security Evaluation in SIEM Systems // International Transac-tions on Systems Science and Applications, Vol.8, December 2012. pp.129-147.

64) Z. Syed, A. Padia, T. Finin, L. Mathews and A. Joshi. UCO: A Unified Cybersecurity Ontology. In: Proceedings of the AAAI Workshop on Artificial Intelligence for Cyber Security, 8 p., 2016.

65) GitHub website. Unified Cybersecurity Ontology. Web: https://github.com/Ebiquity/Unified-Cybersecurity-Ontology

66) Flegel U. Pseudonymizing Unix Log Files // Infrastructure Security. vol. 2437 of the series Lecture Notes in Computer Science. 2002. pp. 162–179.

67) Pang R., Paxson V. A high-level programming environment for packet trace anonymization and transformation // Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications. 2003. pp. 339–351.

68) Kotenko I.V., Chechulin A.A. Computer Attack Modeling and Security Evaluation based on Attack Graphs // Proceedings of 7th International Conference on “Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications” (IDAACS'2013). 2013. pp. 614-619.

69) Kotenko I.V., Chechulin A.A. Fast Network Attack Modeling and Security Evaluation based on Attack Graphs // Journal of Cyber Security and Mobility. 2014. vol. 3. no. 1. pp. 27–46.

70) Котенко И.В, Степашкин М.В., Дойникова Е.В. Анализ защищенности ав-томатизированных систем с учетом социо-инженерных атак // Проблемы информационной безопасности. Компьютерные системы. 2011. № 3. С. 40–57.

71) Ghorbani A.A., Lu W., Tavallaee M. Network Intrusion Detection and Prevention // Springer, 2010. 224 p.

72) Файзуллин Р. Р., Васильев В. И. Метод оценки защищенности сети передачи данных в системе мониторинга и управления событиями информационной безопасности на основе нечеткой логики // Вестник УГАТУ. 2013. Том 17. №2(55). С. 150–156.

73) Орлик С. Введение в программную инженерию и управление жизненным циклом ПО. Программная инженерия. Программные требования [Текст] / С. Орлик, Ю. Булуй. – 2004–2005.

74) Windows Security Log Events, available at: https://www. ultimatewindowssecurity.com/securitylog/encyclopedia/Default.aspx/ [дата обращения: 30.05.2018]