DOI:

МОДЕЛИ НОРМАЛИЗАЦИИ ДАННЫХ В СИСТЕМАХ УПРАВЛЕНИЯ СОБЫТИЯМИ БЕЗОПАСНОСТИ РТК

Исхаков А.Ю., Исхаков С.Ю.

Институт проблем управления им. В.А. Трапезникова РАН, Россия, г. Москва
ул. Профсоюзная д.65

iay@ipu.ru, iskhakov.sy@gmail.com

Аннотация: Рассмотрены проблемы нормализации данных в системах управления событиями безопасности РТК и их негативное влияние на развитие методов корреляции в подобных комплексах. Предложено методическое обеспечение для построения правил корреляции и модифицированная модель жизненного цикла инцидента с учетом источников событий.

Ключевые слова: безопасность, нормализация, корреляция, жизненный цикл инцидента, робототехнические комплексы.

Введение

Применение методов группового управления роботами показало свою эффективность в различных сферах жизни – от бытового применения до промышленных задач и специальных операций. Это обусловлено не только возможностью расширения функционала за счет гетерогенности используемых программно-аппаратных комплексов, распараллеливания задач и рассредоточения мобильных агентов по всей рабочей зоне. Зачастую данный эффект достигается за счет возможности перераспределения целей между роботами группы в том числе в случае выхода из строя некоторых из них. Несмотря на популярность робототехнических комплексов, способных выполнять поставленные задачи с минимальным привлечением человека-оператора, а также повсеместного распространения сетецентрических систем управления, одной из важнейших задач при обеспечении надлежащего уровня защищенности робототехнических групп является автоматизированный мониторинг событий информационной безопасности как центра управления, так и всей инфраструктуры. Целью подобных решений является своевременное оповещение и предупреждение об обнаруженных аномалиях в работе системы, прямо или косвенно являющихся свидетельством инцидента безопасности.

Однако, зачастую подобные оповещения характеризуются большим количеством ошибок первого рода, наличие которых может зависеть от расположения сенсоров, структуры источников данных о событиях безопасности и возможности применения расширенной логики правил на базе сложных организационных иерархий, таких как учетные записи пользователей, критичность контролируемых активов, уровни риска и т.д. Более того, отдельные устройства, включая робототехнические системы, могут быть подвержены эксплуатации злоумышленниками даже в процессе их контроля, поскольку современные атаки зачастую выполняются с помощью стандартных легитимных инструментов, используемых при обслуживании целевой системы. Такие ограничения приводят к множеству оповещений, нарушающих работу аналитиков и дальнейшему возникновению ошибок второго рода из-за пропуска реальных свидетельств действий злоумышленников.

В последнее время ведущие вендоры в области защиты информации выпустили на рынок целый спектр программных и программно-аппаратных решений, позволяющих осуществлять управление событиями безопасности, получаемыми из подсистем журналирования различных источников. Подобные разработки относятся к классу систем управления событиями безопасности (security information event management, SIEM) и позволяют проводить комплексный анализ регистрируемых событий с точки зрения безопасности в контексте онтологий различных этапов атаки, что обуславливает их аддитивную ценность в обнаружении и предотвращении угроз. Среди основных задач таких систем можно выделить:

- - сбор данных и приведение их к единому нормализованному виду;
- - группирование данных по определенным признакам и атрибутам;
- - выявление инцидентов на основе обнаружения корреляций и оповещение персонала служб безопасности;
- - визуализация обрабатываемых данных как инструмент анализа и проведения расследования инцидентов;
- - создание отчетов о состоянии активов защищаемой системы.

Помимо коммерческих разработок в данной отрасли активно ведутся научные исследования, ориентированные на создание методологических основ и поиска новых вычислительных методов для

решения подобных задач. Например, в работах [1, 2] предложен подход к формированию экспертных знаний при разработке SIEM-систем. В основу подхода авторами положена приоритезация определенных признаков и эвристическое выявление скрытых несовместимостей между компонентами защиты объекта. В предлагаемом подходе используются корреляции событий на основе шаблонов, что позволяет обеспечить требуемую в рамках конкретной задачи точность выявления аномалий и при необходимости расширять списки шаблонов. Тем не менее, в работах не учитывается взаимное влияние контролируемых активов и сторонние факторы, влияющие на состояние защищенности устройств. В [3] представлена система признаков для построения модели события в SIEM, а в [4] авторами предлагается методика визуализации метрик кибербезопасности, как инструмент поддержки принятия решений в процессе анализа рисков. Схожее исследование с элементами визуализации данных представлено в [5]. Однако, основными направлениями исследований в этой области научного знания являются методы и механизмы нормализации, агрегации и корреляции событий. В данной статье будут рассмотрены и структурированы актуальные проблемы трансформации данных на различных этапах нормализации при обработке потоков событий в системах управления инцидентами безопасности, а также методологические основы построения правил корреляции.

1 Постановка проблемы

1.1 Проблема потери данных при нормализации

Как упоминалось ранее, цель применения SIEM-систем состоит в обеспечении непрерывного мониторинга информационной безопасности инфраструктуры. Причем для событий, собранных с разных источников, могут отличаться не только формат (TXT, XML, JSON и др.) [2], но и стандарт записи. Для анализа потока событий необходимо привести все события к единой форме, т.е. провести процедуру анализа содержащейся в исходном событии информации в соответствии с заранее заданной для источника и типа события формулой нормализации. Нормализованное событие представляет собой совокупность некоторых полей, состав которых определен в рамках некоторой таксономии, заполненных данными из необработанного события согласно правилам, указанным в формуле нормализации. При этом нормализация полученных событий подразумевает анализ данных исходного события и сопоставление изъятых данных таксономическим полям. Между источником данных и SIEM находятся коннекторы, представляющие некоторый скрипт или приложение, которое позволяет получать журналы событий с источника и проводить процедуру нормализации.

Проблема потери данных при обработке в системах управления событиями безопасности (SIEM) связана с несколькими этапами «упрощения» модели исходного объекта или процесса. Рассмотрим запуск и выполнение отдельно взятого процесса в операционной системе (ОС) (рисунок 1). Процесс занимает определенную часть оперативной памяти и начинает выполнять какой-то набор инструкций. ОС может передавать данные о его протекании в локальный журнал событий, при этом набор регистрируемых данных зависит от настроек логирования. Даже при включении максимального уровня аудита часть информации о протекающем процессе может не попасть в журнал, потому что перечень данных регламентирован при разработке операционной системы. Это приводит к потере некоторых деталей и упрощению исходной модели процесса.

После поступления данных в модуль логирования их необходимо сохранить в журнал, где в качестве ограничителей могут выступать формат хранения записей или протокол, используемый для передачи информации на внешний сервер, где четко определены схема и размеры полей данных. Вероятно, что на этом этапе произойдет очередное «упрощение» модели и часть информации также будет потеряна, поскольку данные, не уместившиеся в эту схему, будут отброшены.

В результате для обработки в SIEM поступает значительно упрощенная модель реального события в виде дискретной записи или структурной единицы данных. Во время нормализации событий полученная информация распределяется по определенным полям. При этом очевидно, что на этом этапе процесса количество полей не полностью отражать все возможные варианты семантических данных для любого из источников. Здесь в очередной раз происходит исключение части информации о процессе или объекте, что дополнительно «упрощает» модель. Поэтому существуют такие поля, куда при нормализации помещают данные, которые не удается отнести к остальным значениям таксономии. Зачастую здесь хранится большой объем разнородной по смыслу информации. Такое часто встречается в SIEM с жестко фиксированным набором таксономических полей. В то же время большой перечень полей при нормализации однозначно приведет к дублированию семантических значений, когда одни и те же данные могут быть определены в различные поля нормализованного события. Подобная

ситуация характерна для SIEM, в которых реализована возможность добавления полей на этапе нормализации, т.е «расширение» модели.

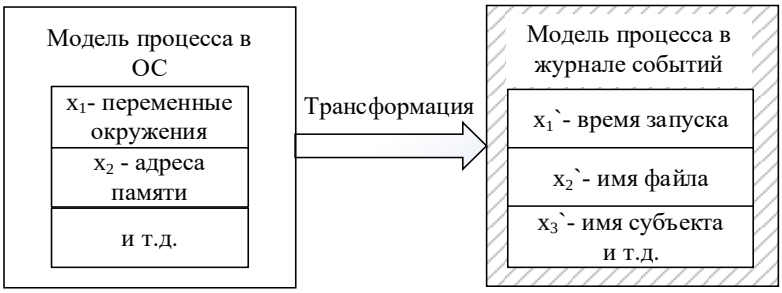

Рис. 1. Потеря данных на уровне ОС

Следствием обоих вышеприведенных случаев являются ситуации, когда при обработке события потеряна актуальная для реагирования на инцидент информация или наоборот имеющиеся данные противоречат друг другу. Вследствие этого оператору или эксперту, проводящему расследование, необходимо либо обратиться к исходному событию, либо моделировать ситуацию, опираясь на собственный опыт.

2.2 Проблемы построения правил корреляции

Широкая вариативность сообщений, генерируемых источниками событий, обуславливает необходимость осуществлять фильтрацию и укрупнение выборок данных, подлежащих анализу, с применением методов агрегации. Поток данных с источника может содержать однотипные события, отличающиеся значением одного или нескольких полей (например, временем регистрации). В подобных случаях целесообразно организовать процесс отбора по условиям заранее определенных правил агрегации данных. При этом несколько событий объединяются в одно агрегированное событие. Имея в распоряжении нормализованные и агрегированные данные о состоянии информационной безопасности инфраструктуры, можно осуществлять действия, направленные на анализ происходящих в системе процессов и выявление инцидентов. Основным механизмом в решении данной задачи является применение методов корреляции.

Выявление корреляции событий – процесс обнаружения инцидентов информационной безопасности путем анализа потока нормализованных событий. При обнаружении в потоке событий такой последовательности, которая указана в условии одного из заранее настроенных правил корреляции, регистрируется корреляционное событие. Среди известных методов корреляции можно выделить два класса – сигнатурные и бессигнатурные. Сигнатурные методы подразумевают создание человеком неких правил определения инцидентов. Бессигнатурные основаны на обнаружении аномалий по принципу черного ящика, среди которых выделяются подходы, основанные на спецификации, и базирующиеся на интеллектуальном анализе данных.

Методы корреляции применялись в рамках систем обнаружения вторжений для выявления связей между сетевыми событиями с целью их агрегации и последующего обнаружения атак [6-8]. Однако корреляция событий, происходящих в информационной системе, кроме выявления и прогнозирования инцидентов, может применяться для решения различных задач безопасности, в том числе для определения взаимосвязей между разнородными сервисными данными роботов, для группировки низкоуровневых событий, выявления типов объектов информационной системы и связей между ними. Широко распространено использование методов корреляции на основе правил. Основным недостатком здесь являются временные затраты на разработку и модификацию правил администратором. Особенно важно, что эффективность этого процесса напрямую связана с квалификацией администратора. Многие методы, в том числе основанные на шаблонах (сценариях), графах, конечных автоматах, и другие, используют различные модели для отображения событий и установления связей между ними, но могут быть реализованы с использованием правил. Среди перспективных направлений в этой отрасли можно выделить создание подходов к корреляции, основанных на самообучении, таких как байесовские сети, вероятностное исчисление событий, иммунные сети, искусственные нейронные сети и другие. Преимуществом здесь является возможность независимой корреляции разнородных событий при минимальном использовании «ручных» операций. Однако, при использовании таких методов необходимо организовать предварительный анализ данных. Эта операция, в свою очередь,

характеризуется сложностью с точки зрения автоматизации. Также возникают сложности при оценке адекватности и качества моделей, а также релевантности обучающей выборки.

В большинстве случаев выявление причины необходимо для адекватной реакции на инцидент или осуществляется в контексте инцидента. При этом, наступлению каждого инцидента предшествуют различные события: сканирование сетевых ресурсов, ошибки авторизации, попытки установить соединение по закрытым портам и т.д. В качестве примера приведем выдержку из проведенного компанией TrendMicro систем защиты промышленных роботов. Так для примера возможности удаленных атак с помощью поисковой системы Shodan был исследован ряд eWon устройства от компании ABB, доступных из сети Интернет. Далее, используя известные уязвимости, исследователям удалось получить доступ к одному из используемых промышленных роботов. На рисунке 2 представлены выдержки из логов, отражающих подключение посредством сервиса APN.

Рис.2. Сервисный бокс ABB (промышленный маршрутизатор eWON с ребрендингом), предоставляющий возможность удаленного подключения к роботу

Очевидно, что в журналах данного устройства было зафиксировано множество событий при попытках авторизации. Правило корреляции позволяет объединить и сгруппировать подобные предшествующие события для определения момента возникновению инцидента. Правило имеет триггер, срабатывающий по условию, счетчик и сценарий реакции. Часть систем включают интуитивно понятные правила в графическом режиме. Счетчики используются для подсчета количества совпадений по одному и тому же правилу. Правила корреляции могут включать условия различной сложности. Одним из наиболее практичных подходов является применение правил, ограничивающих глубину корреляции и разделение базы событий на онлайн и архивную части.

Рассмотренные выше недостатки современных научных работ и коммерческих продуктов в области управления событиями безопасности приводят к низкому уровню работоспособности правил корреляции на практике и необходимости трудоемкой неавтоматизированной доработки подобных комплексов на каждом конкретном объекте. Совокупность ограничивающих факторов может быть охарактеризована как отсутствие методологических основ для разработки правил корреляции. Тем не менее, сформулированные выше недостатки свидетельствуют о невозможности формирования подобных методологий по разработке правил корреляции событий в отрыве от основных проблем в методическом и алгоритмическом обеспечении процесса управления инцидентами.

Еще одной немаловажной проблемой являются разногласия при создании схем нормализации данных, поскольку эту задачу зачастую решает большое количество экспертов на разных этапах интеграции. Например, во многих продуктах пользователи и представители компаний-интеграторов могут самостоятельно добавлять новые источники и осуществлять нормализацию событий. При этом каждый эксперт может по-разному определить смысловую нагрузку некоторых данных, что приведет

к расположению семантически схожих данных в различных полях нормализации. Это характерно даже для таких общепринятых понятий как ip-адрес инициатора и ip-адрес цели взаимодействия – например, в случаях обмена данными в рамках одной установленной TCP-сессии. В некоторых из перечисленных выше исследований предпринята попытка формализовать процесс нормализации, однако актуальной является необходимость создания четких методологических указаний по нормализации событий от различных источников, в рамках которой было бы явно указано:

1. Какие поля должны подлежать анализу.
2. Какие типы данных соответствуют этим полям.
3. Какая информация актуальна для событий каждого типа.
4. Каковы правила заполнения полей.

Другим немаловажным упущением вышеперечисленных исследований является недостаточная проработанность вопроса контроля изменения состояний защищаемых автоматизированных систем, включая ретроспективный анализ состояния активов. В процессе функционирования систем неизбежно происходят изменения в их архитектуре и структуре как результат действий администраторов, пользователей и другого персонала, взаимодействующего с объектами. Активы системы (объекты) постоянно переходят в различные состояния, что приводит к необходимости поддержки регулярного обновления модели объекта защиты. Таким образом, можно выделить следующие актуальные проблемы в исследуемой предметной области:

- потеря данных при нормализации, связанная с трансформациями моделей.
- отсутствие методического аппарата нормализации событий.
- отсутствие поддержки изменений модели объекта защиты.
- отсутствие методологии написания правил корреляции.

Некоторые из данных проблем относятся к правильной трактовке схемы события, другие же связаны с недостаточным уровнем современного методического обеспечения.

2 Моделирование инцидентов при построении правил корреляции

На сегодняшний день активное развитие получают инструменты выявления атак, в том числе распределенных. В связи с этим отдельно стоят категории источников, генерирующих еще один тип событий – события о выявленных атаках. Сюда можно отнести системы поставки индикаторов компрометации, системы защиты конечных рабочих станций и поведенческого анализа. Подобные решения на реализуют анализ событий двух ранее перечисленных классов и позволяют определить различные этапы атаки на защищаемую инфраструктуру. При этом под атакой понимается последовательность действий злоумышленника, направленных на достижение конкретной цели. Свидетельством атаки может быть одно или несколько событий, прямо или косвенно свидетельствующих о нарушении действующих политик информационной безопасности.

Существует несколько подходов к категоризации атак. Например, база знаний MITRE Adversarial Tactics, Techniques, and Common Knowledge (ATT&CK) [9]. В ней в виде матрицы представлены регулярно обновляемые структурированные описания поведенческих шаблонов злоумышленников. Основным назначением данного инструмента является описание и категоризация техник и методов кибератак, основанные на реальных примерах. В случае с SIEM применение подобных знаний при построении правил корреляции позволяет извлекать из нормализованных событий необходимый и достаточный набор метаданных для выявлений определенной тактики или метода, используемого злоумышленниками. Тем не менее, MITRE ATT&CK не может выступать единственной методологией при разработке контента SIEM, поскольку многие техники встречаются на разных фазах, что может затруднить верное построение правил.

Также известны модели, рассматривающие процесс кибератаки с точки зрения алгоритма действия злоумышленников. Например, модель kill-chain [10]. Этот термин впервые стал широко применяться в отрасли информационной безопасности в начале 2010-х годов для описания последовательности шагов злоумышленника при совершении атаки. Модель Kill-Chain, предложенная компанией Lockheed Martin, аккумулирует ранее существовавшие модели и состоит из 7 взаимосвязанных этапов (рис. 3а): Reconnaissance (Разведка), Weaponization (Вооружение), Delivery (Доставка), Exploitation (Заражение), Installation (Инсталляция), Command and Control (Получение управления), Actions on the Objective (Действия на объекте). Конечно, на сегодняшний день можно сказать, что данная модель не является полностью универсальной. Например, в ней не рассматривается горизонтальная разведка внутри инфраструктуры (lateral reconnaissance) с целью определения более уязвимых систем путем повторения действия, начиная с первой фазы.

Также известна модель жизненного цикла атаки, предложенная компанией Mandiant [9], включающая в себя 8 фаз, в том числе итеративный процесс «закрепления» злоумышленников внутри инфраструктуры. Модель рассматривает возможность ветвления и рекурсии на стадии внутренней разведки. При этом этапы вооружения, доставки зловредного ПО и заражения объединены, тем самым упрощая и дополняя исходную модель от Lockheed Martin. Кроме того, ключевым отличием этой модели является выделение этапа повышения привилегий, на котором используются различные инструменты, в том числе легитимные службы и процессы. Многие исследования в области практической информационной безопасности посвящены обнаружению индикаторов и поведенческих сигнатур, отражающих попытки злоумышленников получить доступ к дополнительным ресурсам в скомпрометированной системе.

Совокупность вышеупомянутых моделей может быть использована для разработки методического обеспечения и формирования подходов к классификации данных, обрабатываемых в SIEM-системах с точки зрения выявления конкретных стадий действий злоумышленников вместо количественного анализа больших объемов данных о событиях безопасности с гетерогенных источников. Каждая из моделей Lockheed Martin, Mandiant и MITRE позволяет сформировать структуру, способствующую категоризации нормализованных данных и выстроить их с точки зрения всего жизненного цикла инцидента. Но все эти модели не учитывают распределение типов метаданных для построения правил корреляции. MITRE содержит дополнительную информацию о методах атак, индикаторах компрометации и методах противодействия для каждой фазы. Но перечень фаз не всегда применим для классификации событий при построении правил.

В [9] исследователи Bryant и Saiedian предложили модель (рис. 3б), где предприняты попытки объединить рассмотренные выше модели с данными из матрицы MITRE. Однако, предложенный подход к построению онтологии SIEM, также не является универсальным и имеет ограничения в части зависимости от таксономии, заложенной в систему производителем.

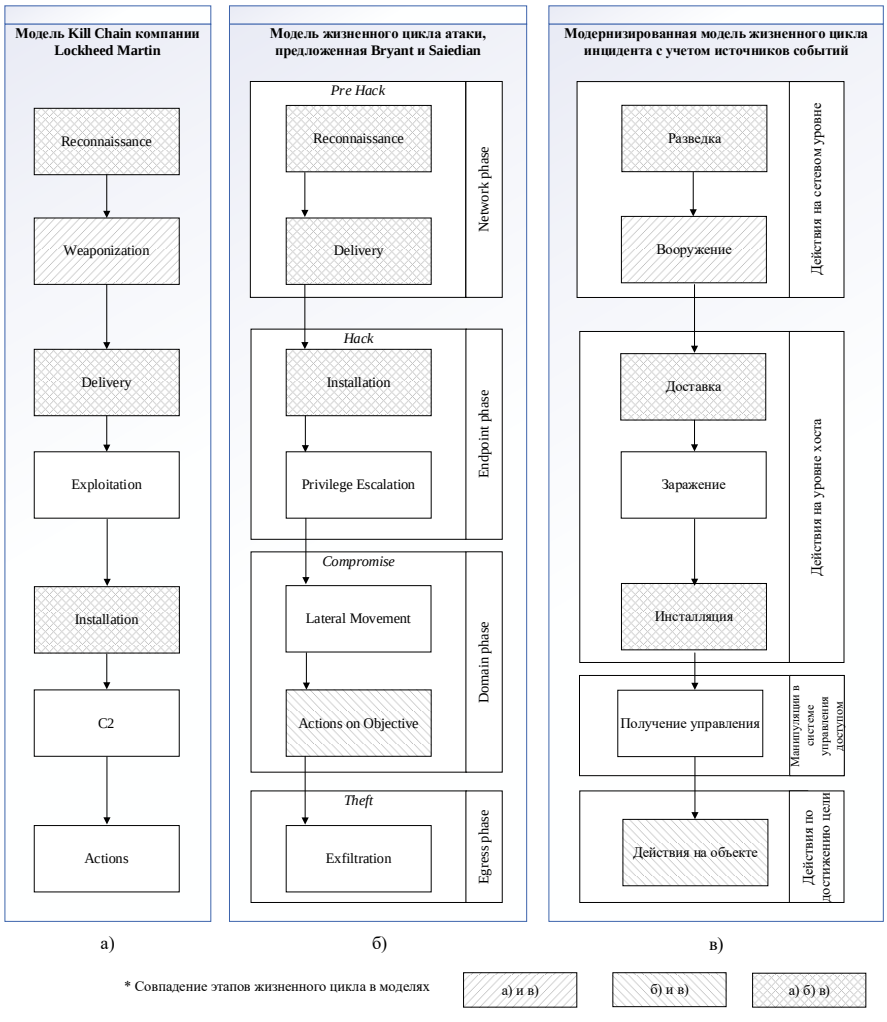

Рис. 3. Сравнение моделей жизненного цикла инцидента

Поскольку в рамках данного исследования авторами для проведения эксперимента использовался стенд на базе системы управления событиями безопасности Arcsight, то на базе модели Kill-chain и модели Bryant и Saiedian была построена модель жизненного цикла инцидента с точки зрения злоумышленника и оптимизирована для использования с данным инструментом (рисунок 3в). При этом учтены принципиально возможные категории источников событий для выявления каждого из этапов. Оптимизированная модель была использована для создания набора правил корреляции и необходимого для них перечня объектов (фильтры, активные листы, мониторы данных и т.д.) с учетом нормализации имеющегося набора источников на основе предложенной выше методологии. В таблице 1 агрегирована информация о категориях возможных источников событий и типах метаданных, извлекаемых из нормализованных событий, которые могут использоваться как индикаторы компрометации или условия при построении правил корреляции.

Таблица 1. Категории источников событий и примеры метаданных для формирования индикаторов компрометации в соответствии с фазами жизненного цикла инцидента

Фаза жизненного цикла инцидента	Возможные источники	Индикаторы для построения правил
Разведка	Межсетевые экраны, сетевые системы обнаружения вторжений	IP-адреса, DNS-имена, номера портов
Вооружение	Сетевые системы обнаружения вторжений	Сигнатуры поведенческих аномалий
Доставка	Контроллеры домена, почтовые сервера, прокси-сервера	Записи об ошибках входа, вложения в письма, скачиваемый контент
Заражение	Хостовые антивирусы, системы обнаружения вторжений уровня хоста	URL, хеши файлов, темы писем, отправители писем
Инсталляция	Контроллеры домена, журналы рабочих станций, системы обнаружения вторжений уровня хоста	Имена пользователей, имена файлов и директории
Получение управления	Межсетевые экраны, хостовые антивирусы	IP-адреса, DNS-имена, номера портов
Действия на объекте	Межсетевые экраны, сетевые системы обнаружения вторжений	IP-адреса, DNS-имена, номера портов, статистика трафика

3 Проведение эксперимента

Для проведения экспериментов использовался кластер из двух нод на платформе Supermicro с двумя процессорами Intel Xeon E5 2.1 ГГц и объемом ОЗУ 128 Гб. Были созданы виртуальные машины на базе операционных систем Windows Server 2016 и Windows 10 для построения доменной структуры, а также CentOS для развертывания ArcSight, используемый для сбора и обработки данных журналов. В качестве активного сетевого оборудования использовались коммутаторы CISCO Catalyst, точки доступа Mikrotik и межсетевой экран CISCO ASA. Беспроводной сегмент использовался для включения в лабораторную сеть робототехнических комплексов. На рабочих станциях было развернуто специализированное программное обеспечение, позволяющее управлять этими РТК. В свою очередь на серверах были развернуты экземпляры баз данных для сбора показаний системы. В качестве входных данных использовался поток событий по протоколу syslog с сетевого оборудования, журналы событий ОС, а также данные из СУБД MS SQL Server 2016. Нормализация специализированных источников событий в виде модулей РТК не проводилась, данная задача является заделом для других этапов исследования. Интенсивность общего потока событий, усредненная по часам за сутки исследования, составила около 600 событий в секунду (events per second, EPS).

Для апробации предложенных подходов было проведено сравнение пакета правил, сформированного из набора, предоставляемого вендором, а также свободно распространяемого контента на платформе Microfocus Marketplace. На втором этапе эксперимента был протестирован набор правил и связанного с ним, контента, предложенные авторами и созданного в соответствии с модернизированной моделью жизненного цикла инцидента. Обычно при разработке контента SIEM наиболее распространено создание правил, позволяющих детектировать хорошо известные

поведенческие сигнатуры действий злоумышленников, отражающихся в журналах событий. Для этого необходимо регулярно использовать материалы вендора и специализированные ресурсы.

Среди созданного для проведения эксперимента набора правил корреляции можно выделить несколько групп, в которых использовались разные подходы при их формировании. Это правила на базе одной группы условий, на базе нескольких групп условий, а также основанные на статистических наблюдениях.

Правила на базе одной группы условий. Один из простейших примеров, основанный анализе метаданных, содержащихся в нормализованных событиях и их сравнении с перечнями индикаторов компрометации. Например, сравнение полей нормализованных событий с обновляемыми спискам IP-адресов, связанных с распространением вредоносного программного обеспечения.

Правила на базе нескольких групп условий. Основаны на последовательном применении нескольких условий и ориентированы для сортировки событий и получения дополнительного контекста при формировании уведомлений. Однако, такой подход имеет ряд ограничений. Во-первых, применение последовательной схемы выполнения условий снижает вероятность того, что все они будут выполнены, с добавлением каждой новой ступени. С одной стороны, многоступенчатость позволяет снизить количество ошибок первого рода, но с другой стороны на практике есть большая вероятность, что правило не сработает из того, что не было выполнено хотя бы одно условие, в т.ч. в случае неверной конфигурации логирования на источнике. Это приводит к высокой вероятности возникновения ошибок второго рода, т.е. пропуску реальных инцидентов. Во-вторых, подобный подход снижает количество полезной информации для аналитиков при криминалистическом анализе инцидента. Причина этого в том, что при проверке каждой группы условий нормализованные события агрегируются и все метаданные, которые не входят в эти условия могут быть не включены в описание инцидента. Различные вариации с увеличением количества агрегируемых полей и метаданных приведут к усложнению условий и повышению требований к ресурсам для их проверки. В данном случае для получения дополнительных метаданных, необходимых для анализа инцидента, необходимо в неавтоматизированном режиме делать запросы по каждому из нормализованных событий. Таким образом, подобные правила либо содержат мало дополнительных полей для группировки, либо ограничены конкретными источниками данных с известными согласованными полями метаданных для запроса.

Правила на основе статистических наблюдений. Используются для выявления аномалий в работе контролируемых источников. Подобный подход подразумевает построение правил из нескольких модулей, одни из которых формируют основные критерии на основе контроля выбранных параметров, а другие отвечают за установление и контроль превышения пороговых значений. В данном случае, несмотря на гибкость формирования критериев, также теряется часть метаданных, и вероятно снижение производительности за счет хранения в памяти статистических данных для сравнения со вновь поступившими значениями.

4 Результаты и обсуждение

На каждом из наборов правил были проведены наблюдения за срабатыванием оповещений правил корреляции в периоды по тестированию на проникновение с использованием свободно распространяемого дистрибутива Black Arch Linux. При этом перед началом эксперимента была тщательно имитирована работа пользователей в корпоративной сети: создан домен и пользователи, настроено сетевое оборудование и выделены несколько vlan, развернуты базы данных и сетевые папки, организован обмен контентом, в т.ч. за пределы лабораторной сети. После этого были сохранены образы всех виртуальных машин и конфигурации оборудования, которые затем загружались перед тестированием каждого набора. Таким образом удалось добиться идентичности условий функционирования. В таблице 2 представлены количество сгенерированных инцидентов по различным примененным тактикам.

Таблица 2. Сравнение базового набора правил и набора на основе предложенной модели

№	Примененная тактика	Базовый набор	Набор на основе предложенной модели
		Количество срабатываний	Количество срабатываний
1	Сканирование портов Nmap	0	5
2	Поиск уязвимостей сканером Open VAS	10	1

№	Примененная тактика	Базовый набор	Набор на основе предложенной модели
		Количество срабатываний	Количество срабатываний
3	Попытка подбора пароля	82	5
4	Отправка фишинговых писем	0	1
5	Скачивание подозрительных файлов	0	5
6	Сканирование сетевых папок	55	3
7	Попытка передачи данных за пределы периметра	3	5
8	Подбор пароля к RDP	44	6
9	Повышение привилегий до локального администратора	0	9
10	Попытка отключения антивируса	5	1

Анализ полученных результатов показал, что используемый набор контента, основанный на модифицированной модели в большинстве случаев позволяет оптимизировать количество оповещений для операторов, в то время как объем метаданных, включаемый в состав скоррелированных событий не меньше, чем при использовании базового контента. Одной из основных причин такого результата является ориентированность базового набора на конкретные значения нормализованных полей, что в случае активных действий злоумышленника вызывает множество однотипных по содержанию событий в журналах безопасности. В тех случаях, когда в этом наборе использовались приемы агрегации (группировки по некоторым конкретным таксономическим полям), наблюдались факты ошибок второго рода, т.е. «пропуск» реальных свидетельств инцидента безопасности.

Заключение

Активное внедрение в жизнедеятельность человека различных робототехнических комплексов обусловлено возможностями их интеграции на базе существующей сетевой инфраструктуры как в виде классических сетей передачи данных, так и различных платформ интернета вещей. Помимо очевидного преимущества и получения эффективного инструмента для управления множеством гетерогенных устройств это обстоятельство приводит к наследованию ряда существенных уязвимостей. Эксплуатация современных сетей передачи данных связана с новыми методами и тактиками киберпреступлений, а современные системы обнаружения вторжений и управления событиями безопасности либо предоставляют аналитикам недостаточное количество информации о происходящих на объекте событиях, либо генерируют чрезмерное количество оповещений, что приводит к снижению скорости реагирования на инциденты и качества их расследования. Несмотря на значительное количество представленных на рынке SIEM-систем, наблюдается недостаток методического обеспечения процессов нормализации событий, а также построения правил корреляции.

В данной статье авторами была предпринята попытка формализовать проблемы потери данных при трансформации моделей в процессе нормализации событий и существующие модели жизненного цикла современных кибератак. При этом модифицированная авторами модель жизненного цикла инцидента была также положена в основу предложенного методического обеспечения по разработке правил корреляции. На основе предложенных наработок был создан набор правил для SIEM ArcSight, одного из лидеров данного рынка, и проведено сравнение с базовым набором, дополненным свободно распространяемым вендором контентом, не учитывающий выделение этапов кибератак. Наблюдения показали, что предложенная конфигурация позволила достичь сокращения количества ошибок первого рода при генерации оповещений, которые не представляют для аналитиков криминалистической ценности. Кроме того, объем метаданных, предоставляемых для каждого коррелированного события, был значительно увеличен за счет механизма агрегации метаданных, специфичных для каждой фазы иницдента.

Повышение рисков безопасности, к числу которых относятся возможность перехвата информационных сигналов, злоумышленного навязывания инструкций робототехническим системам и нарушение доступности критически важных интеллектуальных устройств, обуславливает острую потребность в оценке применения известных и разработке новых эффективных и теоретически обоснованных решений по автоматизированному мониторингу событий безопасности.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта No 19-01-00767.

Литература

1. *Kotenko I., Doynikova E., Chechulin A.* Security Metrics Based on Attack Graphs for the Olympic Games Scenario // Proceedings of the 22nd Euromicro International Conference on Parallel, Distributed, and Network-Based Processing. 2014. – P.561-568.
2. *Lavrova D.* An approach to developing the SIEM system for the Internet of Things // Automatic Control and Computer Sciences. Vol. 50. 2016. – P.673-681.
3. *Raju B.K., Geethakumari G.* Event correlation in cloud: a forensic perspective // Computing. 2016. Vol. 98, № 11. – P.1203–1224.
4. *Проноза А. А., Чечулин А. А., Котенко* И. В. Математические модели визуализации в SIEM-системах // Труды СПИИРАН, 3(46), С.90-107.
5. *Novikova E., Kotenko I.* Analytical visualization techniques for security information and event management // Proceedings of the 21st Euromicro International Conference on Parallel, Distributed, and Network-Based Processing. 2013. – P. 519–525.
6. *Han Y., Zhu M., Liu C.* A Service-Oriented Approach to Modeling and Reusing Event Correlations // Proceedings of the IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC). 2018. - P. 498-507.
7. *Shameli-Sendi A., Louafi H., He W., Cheriet M*. Dynamic Optimal Countermeasure Selection for Intrusion Response System // Proceedings of the IEEE Transactions on Dependable and Secure Computing. Vol. 15. 2018, №5, – P. 755-770.
8. *Исхаков С.Ю., Исхаков А.Ю., Мещеряков Р.В.* Визуализация больших данных с применением методов корреляции событий информационной безопасности в инфраструктуре интернета вещей // Труды Международной конференции по компьютерной графике и зрению "Графикон". №28 2018. – С.311-314.
9. *Bryant B., Saiedian H.* Improving SIEM Alert Metadata Aggregation with a Novel Kill-Chain Based Classification Model // Computers & Security, Vol. 94. 2020.
10. *Hoffmann R.* Markov Models of Cyber Kill Chains with Iterations // Proceedings in the 2019 International Conference on Military Communications and Information Systems (ICMCIS). 2019, P. 1-6.