УДК 004.056
doi:10.15217/issn1684-8853.2018.1.58

КОРРЕЛЯЦИЯ ИНФОРМАЦИИ В SIEM-СИСТЕМАХ НА ОСНОВЕ ГРАФА СВЯЗЕЙ ТИПОВ СОБЫТИЙ

А. В. Федорченко[а, б], *аспирант, младший научный сотрудник, fedorchenko@comsec.spb.ru*
И. В. Котенко[а, б], *доктор техн. наук, главный научный сотрудник, ivkote@comsec.spb.ru*
[а]*Санкт-Петербургский институт информатики и автоматизации РАН, 14-я линия В. О., 39, Санкт-Петербург, 199178, РФ*
[б]*Университет ИТМО, Кронверкский пр., 49, Санкт-Петербург, 197101, РФ*

Постановка проблемы: *в настоящее время системы управления информацией и событиями безопасности являются активно развивающимся и широко распространенным классом средств мониторинга безопасности различных инфраструктур. Неотъемлемым процессом, реализуемым системами данного класса, является корреляция информации для выявления предупреждений и инцидентов безопасности. С учетом роста различных видов источников исходных данных, а также их количества и сложности взаимосвязей между ними существующие подходы не в состоянии обеспечивать эффективное выполнение процесса корреляции.* ***Цель исследования:*** *разработка методики корреляции событий безопасности с автоматизированной адаптацией к анализируемой инфраструктуре, а также создание модели анализа событий безопасности на основе их типов.* ***Результаты:*** *разработана модель корреляции для выполнения структурного анализа входных данных, на основе которой производится построение графа типов событий с прямыми и косвенными связями между ними. Сформулированы требования к нормализации исходных данных по наличию равнозначных свойств в форматах типов событий, а также к полноте и временной синхронизации журналов. Приведен пример анализа журнала событий безопасности, а также полученный в результате граф связей типов событий.* ***Практическая значимость:*** *предлагаемый подход основан на учете различных свойств типов отношений и связей между ними и позволяет использовать ранее не применяемый метод ранговой корреляции наряду с другими методами интеллектуальной обработки информации, что обеспечивает выполнение процесса корреляции событий и информации безопасности с возможностью адаптации к инфраструктуре.*

Ключевые слова *— защита информации, системы мониторинга и управления безопасностью, корреляция событий безопасности.*

Цитирование: Федорченко А. В., Котенко И. В. Корреляция информации в SIEM-системах на основе графа связей типов событий // Информационно-управляющие системы. 2018. № 1. С. 58–67. doi:10.15217/issn1684-8853.2018.1.58
Citation: Fedorchenko A. V., Kotenko I. V. Correlation of Information in SIEM Systems based on Event Type Relation Graph. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2018, no. 1, pp. 58–67 (In Russian). doi:10.15217/issn1684-8853.2018.1.58

Введение

В настоящее время на мировом рынке присутствует достаточно большое количество разнообразных классов средств по обеспечению безопасности различных инфраструктур. Данные средства направлены на предупреждение, обнаружение и предотвращение компьютерных атак и вредоносной активности, а также мониторинг и управление текущим уровнем безопасности. Одним из классов таких средств являются SIEM-системы (Security Information and Event Management), развивающиеся уже на протяжении более 10 лет [1, 2].

Основная задача SIEM-систем заключается в сборе определенной разнородной информации и выявлении в ней высокоуровневых инцидентов и предупреждений о нарушении безопасности [1–4]. Для достижения указанной цели, как правило, применяются методы нормализации, агрегации, фильтрации и корреляции событий. Однако с ростом сложности проводимых атак (например, целевых атак), защищаемых объектов (например, киберфизических систем) и используемых технологий (например, технологии Интернета вещей) применяемые методы и подходы зачастую не в состоянии обеспечивать должный уровень защищенности. Данная тенденция усугубляется возрастающим объемом данных, обработка которых становится все более затруднительной.

Процесс корреляции данных в средствах защиты класса SIEM играет основополагающую роль. Данный процесс в основном направлен на определение причинно-следственных связей между поступающими на обработку событиями. Он позволяет выполнять обнаружение вредоносной и аномальной активности, определение источника и цели атаки, выявление многошаговых атак и зависит от конкретной реализации [5]. Несмотря на разнообразие методов и подходов, применимых в процессе корреляции, на данный момент наиболее распространенным остается правило-ориентированный метод [6–8].

В данной работе представлен подход к корреляции на основе анализа типов событий безопасности в целях определения связей между ними.

Особенностями данного подхода является использование формируемого графа типов событий с прямыми и косвенными связями между собой для выполнения функционального и поведенческого анализа за счет вычисления частотно-временных характеристик событий, определения причинно-следственных связей, ранжирования событий и построения шаблонов поведения.

В работе рассматривается место, роль и общие принципы процесса корреляции, ставится глобальная задача разработки методики корреляции и частная задача построения модели анализа событий безопасности. Рассматриваются отдельные этапы применения предлагаемого подхода. В том числе раскрывается процесс формирования неориентированного графа отношений типов событий, выполняемого на основе анализа типов событий в пределах одного журнала событий, с учетом его нормализованного представления. Рассматриваются особенности использования полученных данных для проведения функционального и поведенческого анализа. Для корректной работы предлагаемого подхода формулируются необходимые и достаточные условия использования исходных (входных) данных. В заключении описывается эксперимент по анализу типов событий журнала безопасности операционной системы (ОС) Windows. Приводятся результаты эксперимента и оценка возможностей применения предлагаемого подхода для корреляции событий безопасности.

Релевантные работы

Корреляция данных первоначально была применена в средствах обнаружения вторжений (Intrusion Detection Systems) для выявления связей между сетевыми событиями с целью их агрегации и последующего детектирования атак (в том числе распределенных и многошаговых) [5]. Именно из систем данного класса методы корреляции были адаптированы для корреляции информации в SIEM-системах.

При общем рассмотрении процесс корреляции может быть разделен на следующие этапы [5]: 1) нормализация; 2) агрегация; 3) фильтрация; 4) анонимизация; 5) приоритезация и 6) собственно корреляция. Наличие и дополнительная декомпозиция данных этапов зависит от конкретной реализации. С нашей точки зрения, каждый из указанных этапов необходим для полноценного выполнения процесса корреляции.

На данный момент существует множество методов корреляции событий безопасности со своими достоинствами и недостатками. На конкретном этапе процесса корреляции целесообразно применять методы, в наибольшей степени удовлетворяющие его задачам в текущей инфраструктуре. Методики выполнения общего процесса корреляции в существующих решениях, как правило, комбинируют несколько методов. Все методы можно условно разделить на сигнатурные и эвристические. Данные методы могут применять различные алгоритмы, основанные на анализе схожести, статистическом анализе, интеллектуальном анализе данных и др.

Сложность оценки качества используемых методик корреляции данных заключается в том, что производители SIEM-систем в целях защиты интеллектуальной собственности не разглашают особенности технологических решений, применяемых в своих системах. К тому же даже при покупке SIEM-системы исследование модуля корреляции затруднено тем, что его настройка заключается преимущественно в формировании новых (дополнительных) правил и исключений.

Однако наряду с платными решениями SIEM-систем существуют также проекты с открытым исходным кодом, а также множество научных публикаций по методам и подходам к корреляции событий безопасности.

Наиболее популярным и простым при реализации является правило-ориентированный метод [6–8], основанный на фиксированном соотнесении событий друг с другом при выполнении определенных условий. Данные условия могут содержать логические операции над данными, их свойствами и вычисляемыми показателями. Главным недостатком данного метода является сложность и длительность составления правил администратором безопасности. Причем эффективность выполнения корреляции правило-ориентированным методом напрямую зависит от квалификации специалиста по внедрению.

Многие методы, такие как шаблонно-ориентированный (сценарно-ориентированный) [6], граф-ориентированный [9, 10], на основе машины конечных состояний [9, 11], на основе схожести [12, 13] и другие, по своей сути имеют различные модели представления событий и их связей, однако в конечном счете они также могут быть выражены в виде правил.

Современным направлением развития методов корреляции событий является применение методов интеллектуального анализа данных, таких как байесовские сети [6, 9, 14], иммунные сети [9, 14], искусственные нейронные сети [9, 14–16] и др. Достоинство данных подходов заключается в возможности самостоятельной (безусловной) корреляции событий с минимизацией ручной настройки. Однако для построения предсказательных моделей на основе интеллектуальных методов требуется предварительный анализ самих данных, который далеко не всегда можно автоматизировать. Кроме того, применение интеллектуальных

подходов определяет требование по оценке адекватности и качества моделей, а исходные данные обучения должны быть достаточно полными.

Подход к корреляции событий

Процесс корреляции

Для постановки задачи исследований следует определить место и роль процесса корреляции в SIEM-системах. Считается, что процесс корреляции направлен на решение следующих задач:

1) группировку низкоуровневых событий в более высокоуровневые события;

2) определение взаимосвязей между разноуровневыми событиями и информацией безопасности;

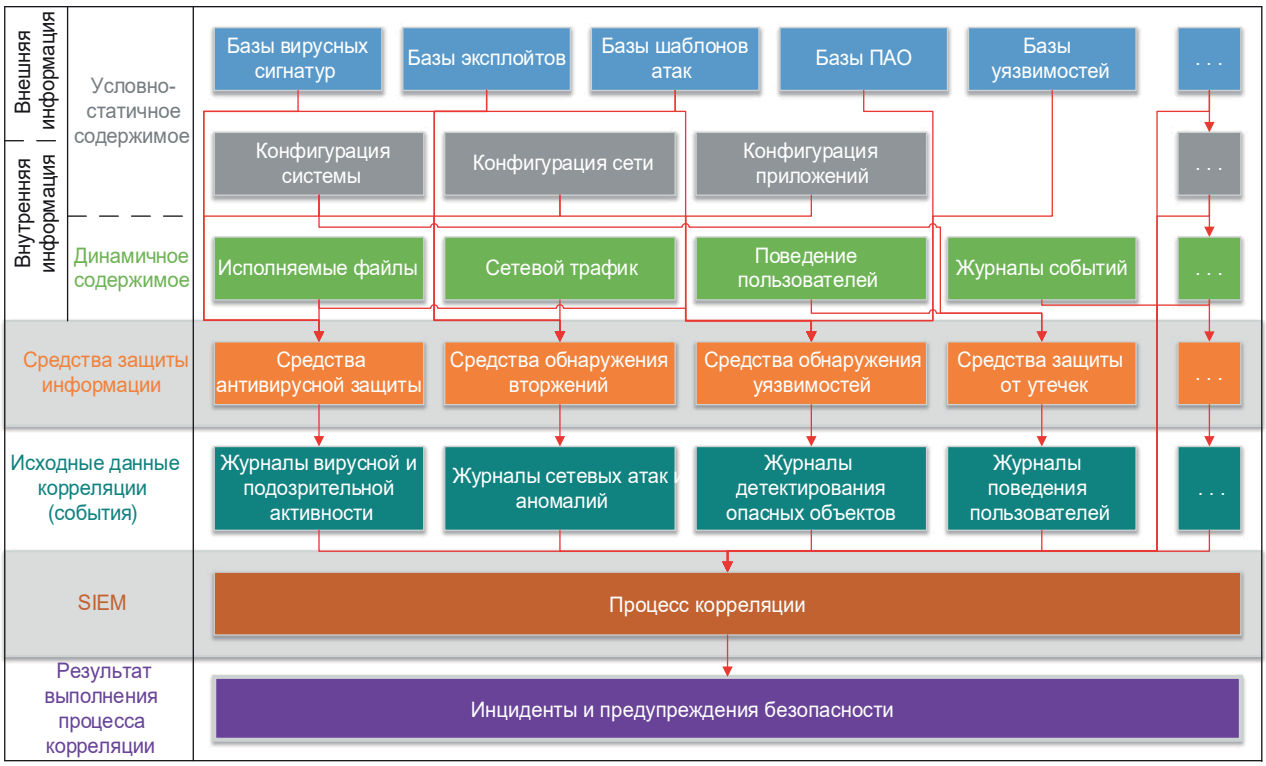
3) определение важности событий и их групп в рамках задачи обеспечения безопасности;

4) обнаружение предупреждений и инцидентов безопасности.

Таким образом, выполнение процесса корреляции начинается со сбора данных из разнородных источников и заканчивается на этапе формирования отчета о текущем состоянии защищенности анализируемых инфраструктур. Стоит также отметить, что процесс корреляции является непрерывным и должен быть рассчитан на выполнение в реальном масштабе времени.

Основная (глобальная) задача проводимых исследований заключается в разработке методики автоматизированной корреляции разнородной информации безопасности. Для достижения данной задачи предлагается использовать результаты структурного, функционального, поведенческого и эволюционного анализа защищаемых объектов [17]. Представленное разделение задачи обусловлено соответствующими аспектами сложности анализируемых инфраструктур как сложных динамических систем [17]. В рамках текущего исследования частной задачей является разработка модели на основе анализа типов событий для выполнения процесса корреляции данных. Новизна предлагаемого подхода заключается в автоматизации поиска прямых и косвенных связей между разнородными событиями для выполнения процесса корреляции. Данный подход должен обеспечить автоматическое включение в модель корреляции типов событий, которые ранее не были известны, но только после приведения данных о событиях к нормализованному виду.

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. В общей схеме



■ ***Рис. 1.*** Общая схема использования входных данных для процесса корреляции в SIEM-системах

■ ***Fig. 1.*** General scheme of using input data for correlation process in SIEM-systems

использования различных источников данных (рис. 1) приведены базы программно-аппаратного обеспечения *ПАО* — это хранилища информации об идентификационных характеристиках установленных программно-аппаратных средств в анализируемой инфраструктуре. 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 ко времени (для динамичного содержимого). Также следует отметить, что на данном этапе разрабатываемый подход ориентирован преимущественно на входные данные с динамичным содержимым, поскольку любое изменение условно-статичной информации может быть также представлено как событие. Однако данный факт не исключает учет данных с условно-статичным содержимым при анализе состояния безопасности. 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, т. е. использование того или иного источника конкретным средством зависит от его реализации. Таким образом, исходными данными для выполнения процесса корреляции являются разнородные и разноуровневые события безопасности, что должно быть обязательно учтено при решении глобальной задачи.

Анализ журналов событий

Под событием понимается результат действия (выполненного, отклоненного, завершенного с ошибкой) или попытки совершения действия, генерируемый источником действия или системой его обработки, имеющий предопределенный формат описания, понятный системе обработки, а также обладающий специфичными свойствами, описывающими само действие.

События различных типов в пределах одного журнала являются исходными данными для проведения исследований и могут быть выражены как

$$\{e_1, e_2, \dots, e_k\} = E_L, \{t_1, t_2, \dots, t_n\} = T_L,$$

где E — множество событий журнала L, а T — множество типов событий журнала L.

Анализ типов событий предлагается производить на основе реальных исходных данных (журналов событий). В данном случае исключается вероятность возникновения ошибок, связанных с изменением формата типов, а при наличии такого изменения подобные события будут зафиксированы.

На основе анализа журнала событий производится формирование множества свойств множества типов событий:

$$\{p_1, p_2, \dots, p_m\} = P_T,$$

где P — множество свойств множества типов T.

Свойства событий можно условно разделить на следующие группы:

— *идентификационные* свойства, значения которых для каждого события уникальны в пределах множества событий одного журнала (группы журналов) или системы (например, идентификаторы записей событий);

— свойства *принадлежности*, значения которых указывают на содержание событий в определенных множествах, таких как тип, провайдер, хост;

— *временные* свойства, отражающие значение времени создания, записи, старта, окончания и других временных характеристик действия;

— свойства *аудита*, определяющие результат выполнения действия, который описывает действие как успешное, запрещенное, завершившееся с ошибкой и др.;

— *информационные* свойства, отражающие специфичные характеристики действия, описываемого в событии (является наиболее обширной группой атрибутов).

Таким образом, анализ журнала событий (с целью выявления структур типов и их свойств) можно представить как отображение множества событий в множества типов и свойств типов событий:

$$E \to T \times P.$$

Модель корреляции
на основе анализа типов событий

Поскольку выявленные типы событий определяются свойствами, характеризующими описываемое в событии действие, связи между типами событий с помощью анализа их структур формируются за счет связей между их свойствами. Для определения места структурного анализа в задаче определения связей необходимо ввести классификацию отношений между свойствами типов событий. Выделим *отношения* по равнозначным и неравнозначным свойствам.

Равнозначное свойство p_r — это одинаковое свойство двух различных типов событий t_1 и t_2:

$$\forall\, p_r \in P_T : p_r \in {}_T t_1, p_r \in {}_T t_2, \text{ где } t_1, t_2 \in T.$$

В свою очередь отношения по *неравнозначным свойствам* делятся на однотипные и разнотипные.

Однотипные неравнозначные свойства p_1 и p_2 — это свойства, эквивалентные по типу содержимого:

$$p_1, p_2 \in P_T : p_1 \sim p_2.$$

Разнотипные неравнозначные свойства — это свойства, эквивалентные по значениям содержимого при явной разнице между типами содержимого.

Кроме того, один тип событий t может содержать несколько однотипных и разнотипных неравнозначных свойств p в своей структуре:

$$\forall\{p_1, p_2, \dots, p_s\} \subset P_t, p_1 \sim p_2 \sim \dots \sim p_s, t \in T,$$

где s — количество однотипных или разнотипных неравнозначных свойств типа t.

При последующем анализе наличие у различных типов событий равнозначных свойств будет рассматриваться как *прямая связь* между свойствами типов событий, а наличие однотипных и разнотипных неравнозначных свойств — как однотипные и разнотипные *косвенные связи* соответственно. В рамках структурного анализа рассматриваются только прямые связи между типами событий, тогда как функциональный и поведенческий анализы подразумевают определение косвенных однотипных и разнотипных связей соответственно.

Например, при сравнении структур двух типов событий безопасности для ОС Windows «Завершение процесса» (4689) и «Вызвана привилегированная служба» (4673) одним из равнозначных свойств обоих типов является «ProcessId» (процесс-инициатор), что является прямой связью между указанными типами событий.

Событие типа «Запуск процесса» (4688), помимо свойства «NewProcessID», содержит свойство «CreatorProcessID». Оба свойства инициализируют идентификатор процесса, только в первом случае — дочернего процесса (наследника), а во втором — процесса-инициатора (предка). Указанная связь является косвенной однотипной по типу содержимого (тип — «Идентификатор процесса»), поэтому позволяет проследить функциональные связи между событиями разных типов и, в данном случае, выявлять события рабочих сессий процессов и их иерархии наследования.

Событие типа «Запуск процесса» (4688) содержит также свойство «ProcessName». При рассмотрении свойств «ProcessId» и «ProcessName» типы их содержимого явно отличаются: в первом случае — это «Идентификатор процесса», а во втором — «Имя исполняемого модуля в файловой системе». Однако оба свойства описывают идентификационные характеристики процесса. В первом случае данная характеристика имеет привязку ко времени — идентификатор присваивается системой каждому создаваемому процессу и имеет уникальное случайное значение в текущий момент времени в рамках отдельной сессии процесса (от создания до завершения). Во втором случае идентификационная характеристика является статичной и не имеет привязки ко времени. При дальнейшей корреляции указанных неравнозначных свойств разного типа может быть определена косвенная разнотипная связь.

В результате анализа структур типов событий, в соответствии с предлагаемой моделью корреляции, формируется неориентированный граф связей G, используемый при дальнейшем анализе входных данных и выполнении процесса корреляции:

$$G = (T, P, \varphi), \varphi : P \to T \times T.$$

Следует учитывать, что полученная модель не является окончательной, а наоборот, является первичным представлением отношений между событиями. Далее, после функционального и поведенческого анализа, данная модель корректируется за счет учета силы связей (весов) между типами событий. Вычисление весов ребер полученного графа производится на основе относительных и абсолютных частотных характеристик по принципу схожести значений связующих свойств. В свою очередь направленность связей, выражающая причинно-следственные отношения между типами событий, определяется за счет частотно-временных характеристик с определенной долей вероятности.

Также стоит отметить, что при последующем выделении групп связных событий, а также определении классов объектов со временем жизни и других сущностей будут получены модели последующих (более высоких) уровней. Под классами объектов со временем жизни понимаются высокоуровневые структуры, определяемые с помощью происходящих событий. Например, ОС, как правило, имеют в качестве неделимого оперативного объекта исполняемый поток. Группа потоков формирует процесс. Процесс в свою очередь может определяться одним или несколькими сервисами. Таким образом, более высокоуровневая модель представления событий должна содержать подобные классы объектов и позволять в дальнейшем связывать между собой данные с динамичным и условно-статичным содержимым.

Статистические методы корреляции

Методы корреляции могут подразделяться на параметрические и непараметрические. В связи с тем, что любое событие характеризуется набором качественных и количественных специфич-

ных свойств, применение параметрических методов (ковариации, линейной корреляции) для определения коэффициента корреляции между событиями затруднено. Однако параметрические методы могут быть использованы для определения линейных зависимостей между свойствами и типами событий. В свою очередь при ранжировании входных данных определенным образом возможно использование непараметрических методов ранговой корреляции. Данные методы предлагается использовать для решения следующих задач: 1) определения связей между свойствами внутри типов событий; 2) вычисления силы и направления связей между типами событий; 3) определения связей между группами событий и 4) поиска взаимосвязей между классами объектов. Первые три задачи реализуют функциональный анализ защищаемой инфраструктуры, а последняя задача необходима для мониторинга поведения отдельных объектов и, следовательно, системы в целом.

Одним из способов ранжирования событий для проведения поведенческого анализа является определение силы связей между типами событий, а также между самими экземплярами событий.

Предлагается выделить следующие виды весов связей:

1) удельные веса прямой, косвенной однотипной и косвенной разнотипной связей между типами событий, задающиеся количеством равнозначных, неравнозначных однотипных и неравнозначных разнотипных свойств соответственно;

2) относительные веса связей между экземплярами событий, определяющиеся отношением количества совпадающих значений свойств к соответствующим удельным весам.

В результате анализа выбранного временного окна в пределах анализируемого журнала формируется набор пар значений относительного веса и интервала времени. Частотный анализ полученных наборов между типами событий, а также применение методов ранговой корреляции позволяют определить причинно-следственные отношения между типами событий и между конкретными экземплярами событий.

Выбор конкретного метода корреляции (например, метода Кендалла, метода Спирмена, метода множественной ранговой корреляции) зависит от этапа, на котором он применяется, и возможности применения исходя из входных данных. Он выбирается экспериментальным путем с учетом эффективности использования.

Эволюционный анализ защищаемой инфраструктуры подразумевает выделение классов типовых элементов (хостов, серверов, ОС, сервисов и др.) и обучение модуля корреляции по вычисленным показателям. В качестве «учителя» в данном случае предполагается использовать компоненты защиты информации, обнаруживающие текущие угрозы и подающие их в виде высокоуровневых событий. Ввиду вероятного наличия циклов в направленном графе связей типов, групп типов и классов объектов применение байесовских сетей затруднено. Данный факт обусловлен тем, что цикличность связей элементов графа теоретически не имеет возможности устранения:

— удаление маловероятных цикличных связей может привести к искажению результата корреляции и пропуску аномальных групп событий;

— упрощение невозможно в связи с использованием низкоуровневых и неделимых (элементарных) событий.

Требования к исходным данным

Предлагаемый подход имеет ряд ограничений на входные данные. Так, предполагается, что перед началом выявления структур типов событий в рамках одной модели формат событий является нормализованным. Нормализация структур преимущественно отражается в следующем условии: структура одного типа события t не должна иметь равнозначных свойств p_1 и p_2:

$$\forall\, p_1, p_2 \in P_t : p_1 \neq p_2,\ t \in T.$$

Данное ограничение необходимо для исключения зацикливания на одном событии в ходе использования предлагаемого подхода. Вместе с тем следует соблюдать нормализованный (однозначный) формат записи свойств событий разных типов.

Стоит отметить, что исходные данные также должны удовлетворять необходимому условию по полноте всевозможных типов событий в рамках рассматриваемой модели и достаточному условию по полноте количества разнотипных событий для выполнения структурного, функционального, поведенческого и эволюционного анализа. Также следует учитывать необходимость соблюдения показателей полноты исходных данных по предоставляемой информации. Данный факт обусловлен условием достаточности анализируемой выборки с помощью метода ранговой корреляции, а также обучаемой выборки. Кроме того, в связи с чувствительностью и привязкой предлагаемого подхода к реальному масштабу времени значения временных свойств событий в рамках одной модели должны быть синхронизированы. Таким образом, для корректного применения предлагаемого подхода к журналу, системе, сегменту или инфраструктуре временные показатели событий должны быть синхронизированы в пределах данного журнала, системы, сегмента и инфраструктуры соответственно.

Определение временного окна журнала событий для выполнения поведенческого анализа следует производить также с учетом требования репрезентативности выборки. Поэтому размер выборки предлагается определять на основе:

1) частотно-временного анализа входных типов событий, использование которого обусловлено периодичностью выполнения ряда процессов и задач в различных системах;

2) учета динамики (частоты) изменения значений равнозначных, неравнозначных однотипных и неравнозначных разнотипных свойств, что обусловливается как случайностью (работой пользователя), так и периодичностью (работой системы) описываемых в журналах действий.

Эксперименты и оценка результатов

В настоящее время анализ структур разных типов событий используется, например, средством аналитического анализа данных безопасности Splunk [18]. В этой системе свойства типов событий применяются для нормализации информации после ее загрузки и последующего индексирования для выполнения задач обработки. В данном случае прямые связи, полученные при анализе исходной информации, могут использоваться в запросах обработки, которые в свою очередь составляются на основе экспертных знаний и их ручной корректировки. Для получения косвенных связей необходимо также проводить дополнительный частотный анализ однотипного содержимого. В указанном решении прямые связи только подразумевают их применение, тогда как в предлагаемом авторами статьи подходе прямые связи берутся за основу модели взаимосвязей между событиями.

В рамках проведенных исследований в качестве исходных данных был проанализирован журнал событий безопасности ОС Windows 8 офисного компьютера, не включенного в локальный домен.

Экспериментальный набор данных обладал следующими характеристиками:

— размер журнала ~7 ГБ в формате XML;

— время записи журнала ~1 мес.;

— время обработки журнала ~50 мин;

— количество событий журнала ~6 700 000;

— количество выявленных типов событий — 80 из 418 заявленных в документации [19] для версии данной ОС (типы событий предыдущей версии не учитывались; число экземпляров событий данной версии составило не больше 20);

— количество выявленных свойств — 158, из них 14 — общие (встречаются во всех типах событий), 53 — уникальные (встречаются только в одном типе событий), 89 — смежные (встречаются более чем в одном типе событий).

В результате анализа представленного журнала был сформирован граф прямых связей типов событий (рис. 2). На рисунке видно, что большинство выявленных типов событий имеют достаточно большое количество прямых связей. Всего граф содержит 1309 узлов. Также присутствуют типы событий, не имеющие прямых связей ни с одним другим типом.

Фрагмент графа типов событий ОС MS Windows 8 с учетом вычисленных удельных весов связей между ними представлен на рис. 3. В уточненном графе типов событий сила связи отображается за счет толщины дуг.

В таблице приведены наиболее распространенные смежные свойства типов событий; количество типов отображает число типов событий, обладающих конкретным свойством.

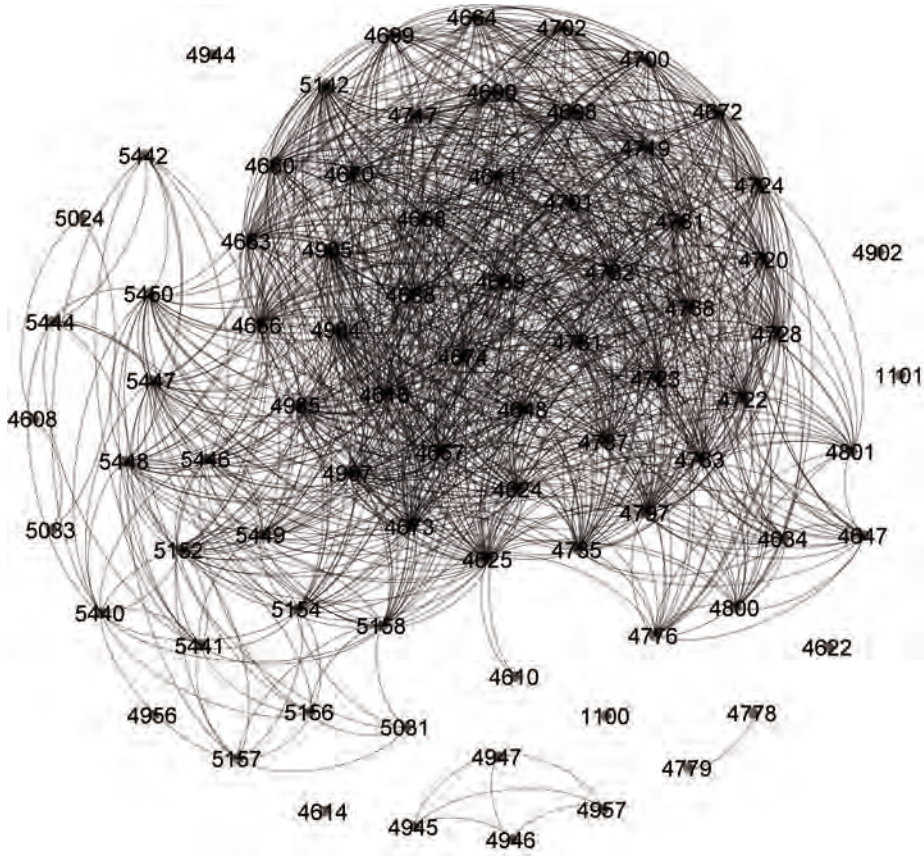
Предлагаемый подход на основе анализа структур типов событий является составной частью общей методики корреляции событий безопасности. Он может быть применен для формирования модели взаимосвязей (корреляции) событий, которая будет уточняться на каждом последующем шаге методики.

Процесс анализа структур типов событий также может рассматриваться как завершающая стадия процесса нормализации. В свою очередь события, коррелируемые в результате работы методики по определенным свойствам их типов, позволят генерировать события более высоко-

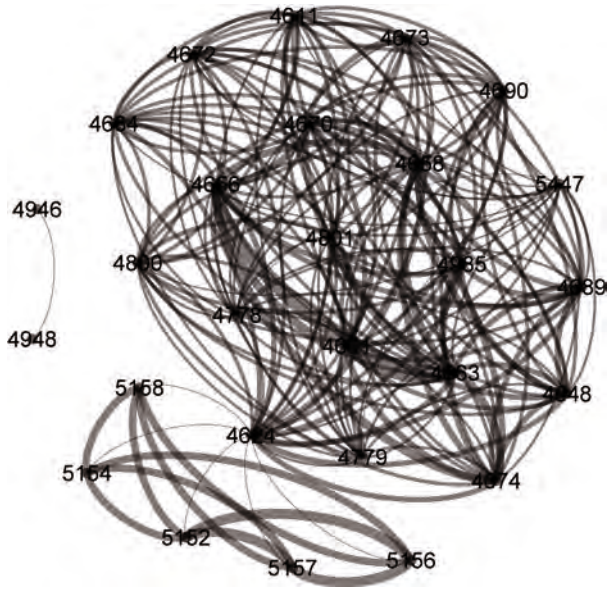
■ Результаты анализа распространенности свойств типов событий

■ Results of the prevalence analysis of event types properties

Свойство	Количество типов	Свойство	Количество типов
SubjectDomainName	28	ProviderKey	9
SubjectLogonId	28	ProviderName	9
SubjectUserName	28	ObjectServer	8
SubjectUserSid	28	TargetUserName	8
ProcessId	25	HandleId	7
ProcessName	16	TargetDomainName	7
LayerName	9	Application	6



■ ***Рис. 2.*** Пример графа связей типов событий
■ ***Fig. 2.*** Example of a events types relations graph



■ ***Рис. 3.*** Пример фрагмента графа типов событий с вычисленными удельными весами связей между ними
■ ***Fig. 3.*** Example of a event types graph fragment with computed the specific weights of the relations between them

го уровня, которые будут содержать составные свойства.

Для обеспечения работы в реальном масштабе времени с учетом требования по полноте исходных данных и теоретически неограниченного размера анализируемой инфраструктуры реализация методики будет нуждаться в применении технологий параллельной и многопоточной обработки больших данных.

Заключение

Проведенные исследования в области корреляции событий безопасности для SIEM-систем показали необходимость разработки новой методики корреляции событий безопасности. На основе анализа общего процесса корреляции и формулировки глобальной задачи в статье предложена модель корреляции на основе анализа типов событий, а также подход к корреляции событий на основе методов ранговой корреляции.

Разрабатываемая методика главным образом направлена на решение таких актуальных задач в области обеспечения безопасности, как монито-

ринг состояния безопасности киберфизических инфраструктур; обнаружение целевых атак (на основе выявления и автоматизированной псевдоклассификации аномалий), а также автоматизированная оценка состояния безопасности инфраструктур неограниченного размера с применением устройств Интернета вещей.

В дальнейшем планируется продолжение разработки общей методики корреляции на основе определения функциональных связей между событиями и построения шаблонов поведения анализируемых инфраструктур.

Работа выполнена при поддержке гранта РНФ № 15-11-30029 в СПИИРАН.

Литература

1. **Kotenko I. V., Chechulin A. A.** A Cyber Attack Modeling and Impact Assessment Framework // Proc. of 5th Intern. Conf. on Cyber Conflict 2013 (CyCon 2013). 2013. P. 119–142.
2. **Kotenko I. V., Polubelova O. V., Saenko I. B.** The Ontological Approach for SIEM Data Repository Implementation // IEEE Intern. Conf. on Green Computing and Communications, Conference on Internet of Things, and Conference on Cyber, Physical and Social Computing. 2012. P. 761–766.
3. **Дойникова Е. В., Котенко И. В.** Методики и программный компонент оценки рисков на основе графов атак для систем управления информацией и событиями безопасности // Информационно-управляющие системы. 2016. № 5. С. 54–65. doi:10.15217/issn1684-8853.2016.5.54
4. **Котенко И. В., Дойникова Е. В.** Методика выбора контрмер на основе комплексной системы показателей защищенности в системах управления информацией и событиями безопасности // Информационно-управляющие системы. 2015. № 3. С. 60–69. doi:10.15217/issn1684-8853.2015.3.60
5. **Kruegel C., Valeur F., Vigna G.** Intrusion Detection and Correlation. Challenges and Solutions. — Springer, 2004. — 118 p.
6. **Sadoddin R., Ghorbani A.** Alert Correlation Survey: Framework and Techniques // Proc. of 2006 Intern. Conf. on Privacy, Security and Trust: Bridge the Gap Between PST Technologies and Business Services (PST`06). 2006. Article no. 37.
7. **Hanemann A., Marcu P.** Algorithm Design and Application of Service-Oriented Event Correlation // 3rd IEEE/IFIP Intern. Workshop on Business-Driven IT Management: Proc. of Conf. BDIM 2008. 2008. P. 61–70.
8. **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.
9. **Muller A.** Event Correlation Engine: Master`s Thesis. — Swiss Federal Institute of Technology Zurich, 2009. — 165 p.
10. **Ning P., Xu D.** Correlation Analysis of Intrusion Alerts // Intrusion Detection Systems. Ser.: Advances in Information Security. Springer, 2008. Vol. 38. P. 65–92.
11. **Ghorbani A. A., Lu W., Tavallaee M.** Network Intrusion Detection and Prevention. — Springer, 2010. — 224 p.
12. **Hasan M.** A Conceptual Framework for Network Management Event Correlation and Filtering Systems // Proc. of the Sixth IFIP/IEEE Intern. Symp. on Integrated Network Management. 1999. P. 233–246.
13. **Zurutuza U., Uribeetxeberria R.** Intrusion Detection Alarm Correlation: A Survey // Proc. of IADAT Intern. Conf. on Telecommunications and Computer Networks. 2004. P. 1–3.
14. **Guerer D. W., Khan I., Ogler R., Keffer R.** An Artificial Intelligence Approach to Network Fault Management. — SRI International, 1996. — 10 p.
15. **Tiffany M.** A Survey of Event Correlation Techniques and Related Topics. http://www.tiffman.com/netman/netman.html (дата обращения: 14.02.2017).
16. **Elshoush H. T., Osman I. M.** Alert Correlation in Collaborative Intelligent Intrusion Detection Systems — a Survey // Applied Soft Computing. 2011. P. 4349–4365.
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. Splunk Security. https://www.splunk.com/en_us/solutions/solution-areas/security-and-fraud.html (дата обращения: 21.06.2017).
19. Windows Security Log Events. https://www.ultimatewindowssecurity.com/securitylog/encyclopedia/Default.aspx (дата обращения: 21.06.2017).

UDC 004.056
doi:10.15217/issn1684-8853.2018.1.58

Correlation of Information in SIEM Systems based on Event Type Relation Graph

Fedorchenko A. V.[a,b], Post-Graduate Student, Junior Researcher, fedorchenko@comsec.spb.ru
Kotenko I. V.[a,b], Dr. Sc., Tech., Chief Researcher, ivkote@comsec.spb.ru
[a]Saint-Petersburg Institute for Informatics and Automation of the RAS, 39, 14 Line, V. O., 199178, Saint-Petersburg, Russian Federation
[b]ITMO University, 49, Kronverkskii Pr., 197101, Saint-Petersburg, Russian Federation

Introduction: Security information and event management systems are now an actively developing and widely distributed class of tools for security monitoring of various computing infrastructures. An essential process implemented by systems of this class is information correlation in order to detect security alarms and incidents. Considering the growth of initial data source types, as well as their number and the complexity of their relationships, the existing approaches cannot provide efficient correlation. **Purpose:** The development of an event correlation technique which automatically adapts to the analyzed infrastructure; in particular, the development of a security event analysis model on the base of event types. **Results:** A correlation model has been developed for structural analysis of the input data. On the basis of this model, a graph of event types is constructed with direct and indirect links between the events. The paper specifies requirements to the initial data normalization concerning the existence of equivalent properties within the formats of the event types, as well as the completeness and temporal synchronization of logs. An example of security events log analysis is provided, along with the resulting graph of the event type links. **Practical relevance:** The proposed approach is based on taking into account various properties of relation types and links between them. It allows you to apply the previously unused method of rank correlation along with other intelligent methods. This provides event correlation and the adaptation to the infrastructure.

Keywords — Information Security, SIEM Systems, Security Event Correlation.

Citation: Fedorchenko A. V., Kotenko I. V. Correlation of Information in SIEM Systems based on Event Type Relation Graph. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2018, no. 1, pp. 58–67 (In Russian). doi:10.15217/issn1684-8853.2018.1.58

References

1. Kotenko I. V., Chechulin A. A. A Cyber Attack Modeling and Impact Assessment Framework. *Proc. of 5th Intern. Conf. on Cyber Conflict 2013 (CyCon 2013)*, 2013, pp. 119–142.
2. Kotenko I. V., Polubelova O. V., Saenko I. B. The Ontological Approach for SIEM Data Repository Implementation. *IEEE Intern. Conf. on Green Computing and Communications, Conference on Internet of Things, and Conf. on Cyber, Physical and Social Computing*, 2012, pp. 761–766.
3. Doynikova E. V., Kotenko I. V. Techniques and Software Tool for Risk Assessment on the Base of Attack Graphs in Information and Security Event Management Systems. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2016, no. 5, pp. 54–65 (In Russian). doi:10.15217/issn1684-8853.2016.5.54
4. Kotenko I. V., Doynikova E. V. Countermeasure Selection in Security Management Systems. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2015, no. 3, pp. 60–69 (In Russian). doi:10.15217/issn1684-8853.2015.3.60
5. Kruegel C., Valeur F., Vigna G. *Intrusion Detection and Correlation. Challenges and Solutions*. Springer, 2004. 118 p.
6. Sadoddin R., Ghorbani A. Alert Correlation Survey: Framework and Techniques. *Proc. of 2006 Intern. Conf. on Privacy, Security and Trust: Bridge the Gap Between PST Technologies and Business Services (PST`06)*, 2006, article no. 37.
7. Hanemann A., Marcu P. Algorithm Design and Application of Service-Oriented Event Correlation. *Proc. of Conf. BDIM 2008, 3rd IEEE/IFIP Intern. Workshop on Business-Driven IT Management*, 2008, pp. 61–70.
8. 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.
9. Muller A. *Event Correlation Engine*. Master`s Thesis. Swiss Federal Institute of Technology Zurich, 2009. 165 p.
10. Ning P., Xu D. Correlation Analysis of Intrusion Alerts. *Intrusion Detection Systems. Series Advances in Information Security*, Springer, 2008, vol. 38, pp. 65–92.
11. Ghorbani A. A., Lu W., Tavallaee M. *Network Intrusion Detection and Prevention*. Springer, 2010. 224 p.
12. Hasan M. A Conceptual Framework for Network Management Event Correlation and Filtering Systems. *Proc. of the Sixth IFIP/IEEE Intern. Symp. on Integrated Network Management*, 1999, pp. 233–246.
13. Zurutuza U., Uribeetxeberria R. Intrusion Detection Alarm Correlation: A Survey. *Proc. of IADAT Intern. Conf. on Telecommunications and Computer Networks*, 2004, pp. 1–3.
14. Guerer D. W., Khan I., Ogler R., Keffer R. *An Artificial Intelligence Approach to Network Fault Management*. SRI International, 1996. 10 p.
15. Tiffany M. *A Survey of Event Correlation Techniques and Related Topics*. Available at: http://www.tiffman.com/netman/netman.html (accessed 14 February 2017).
16. Elshoush H. T., Osman I. M. Alert Correlation in Collaborative Intelligent Intrusion Detection Systems — a Survey. *Applied Soft Computing*, 2011, pp. 4349–4365.
17. Okhtilev M. Y., Sokolov B. V., Yusupov R. M. Intelligent Technologies for Monitoring and Management of Structural Dynamics of Complex Technical Objects. Moscow, Nauka Publ., 2006. 410 p. (In Russian).
18. *Splunk Security*. Available at: https://www.splunk.com/en_us/solutions/solution-areas/security-and-fraud.html (accessed 21 June 2017).
19. *Windows Security Log Events*. Available at: https://www.ultimatewindowssecurity.com/securitylog/encyclopedia/Default.aspx (accessed 21 June 2017).