УДК 004.056
doi:10.15217/issn1684-8853.2017.5.95

ОБЗОР АЛГОРИТМОВ КОРРЕЛЯЦИИ СОБЫТИЙ БЕЗОПАСНОСТИ ДЛЯ ОБЕСПЕЧЕНИЯ БЕЗОПАСНОСТИ ОБЛАЧНЫХ ВЫЧИСЛИТЕЛЬНЫХ СРЕД

Е. С. Новикова[а, б], *канд. техн. наук, доцент*
Я. А. Бекенева[а], *инженер*
А. В. Шоров[а], *канд. техн. наук, ведущий научный сотрудник*
Е. С. Федотов[а], *аспирант*
[а]*Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В. И. Ульянова (Ленина), Санкт-Петербург, РФ*
[б]*Санкт-Петербургский институт информатики и автоматизации РАН, Санкт-Петербург, РФ*

Постановка проблемы: *повышение сложности атак на облачные системы влечет за собой необходимость разработки механизмов защиты, способных анализировать события безопасности, в том числе разнесенные во времени. Кроме того, возникает необходимость в осуществлении проверки достоверности генерируемых событий безопасности, а также сопоставления критичности событий безопасности с уровнем критичности контролируемых ресурсов. В связи с этим необходимо в системе управления информационной безопасностью облачной инфраструктуры использовать модуль корреляции событий безопасности от разных сенсоров безопасности и сетевых устройств инфраструктуры в качестве ключевого компонента системы.* ***Цель:*** *анализ подходов к корреляции событий безопасности для обеспечения безопасности в облачных инфраструктурах.* ***Результаты:*** *анализ основных алгоритмов корреляции событий, а также существующего программного обеспечения, выполняющего корреляцию событий безопасности, показал, что можно выделить три основных подхода к их построению: на основе подобия событий безопасности, на основе знаний и вероятностные. Определены следующие критерии оценки эффективности разработанных методик: возможность анализа данных от разных сенсоров безопасности, требования к наличию предварительных знаний для функционирования модели корреляции, точность корреляции, обнаружение многошаговых и новых типов атак. Представлен сравнительный анализ рассмотренных подходов к корреляции событий безопасности.* ***Практическая значимость:*** *результаты исследований полезны при разработке механизмов защиты облачных вычислительных сред от сетевых атак, в том числе устойчивых целенаправленных угроз. Использование корреляции событий безопасности позволит средствам защиты более точно расставлять приоритеты событиям безопасности и своевременно реагировать на них.*

Ключевые слова *— безопасность облачных технологий, события безопасности, корреляция событий безопасности, алгоритмы корреляции событий.*

Введение

Системы облачных вычислений пользуются большой популярностью как в крупных, так и небольших организациях благодаря легкому конфигурированию вычислительных устройств, гибкости и эластичности предоставляемых облачных сервисов и ощутимой экономической выгоды их использования. Вместе с тем разделение общих вычислительных ресурсов, возможность совместного использования услуг между арендаторами повышает требования к обеспечению безопасного использования информационных ресурсов системы облачных вычислений. Одними из важнейших характеристик таких систем являются их бесперебойное функционирование и защищенность данных пользователей. В связи с этим возрастает необходимость в развитии систем их защиты от различных угроз, включая сложные целевые атаки, выполняемые в несколько этапов, часто разнесенных во времени. Обнаружение таких атак требует тщательного анализа событий безопасности, получаемых от различных датчиков безопасности и объектов облачной инфраструктуры за длительный период времени. Многие современные системы обнаружения вторжений не способны установить взаимосвязи между событиями безопасности в виде последовательности этапов выполнения атаки, так как не имеют инструментов анализа текущих угроз во временном контексте [1]. В большинстве случаев оценка достоверности генерируемых событий безопасности не осуществляется, а критичность события безопасности часто не зависит от уровня критичности контролируемых ресурсов. Корреляция этих двух показателей позволила бы администратору безопасности более точно расставлять приоритеты событиям безопасности для своевременного реагирования на них. Для устранения этих недостатков в системах управления информационной безопасностью предлагается использовать в качестве составного компонента модуль корреляции событий безопасности.

Основными задачами методик корреляции являются:

1) снижение объема исходного потока событий безопасности за счет группирования взаимосвязанных событий, что позволяет снизить когнитивную нагрузку на аналитика;

2) определение взаимосвязей между событиями от разнородных источников, что способствует лучшему пониманию развития атаки в информационной системе;

3) корреляция событий в контексте системы, что позволяет лучше понять сценарий атаки, ее цели и задачи [2]. Существуют различные схемы классификации разработанных методик корреляции событий, незначительно отличающиеся друг от друга [1–3]. В целом можно выделить три основные группы исходя из их особенностей реализации и решаемых с их помощью задач корреляции:

— на основе подобия (сходства) событий безопасности;

— на основе знаний;

— вероятностные (или статистические).

Следует отметить, что представленная классификация алгоритмов корреляции данных не является достаточно точной, поскольку некоторые алгоритмы могут быть отнесены сразу в несколько категорий.

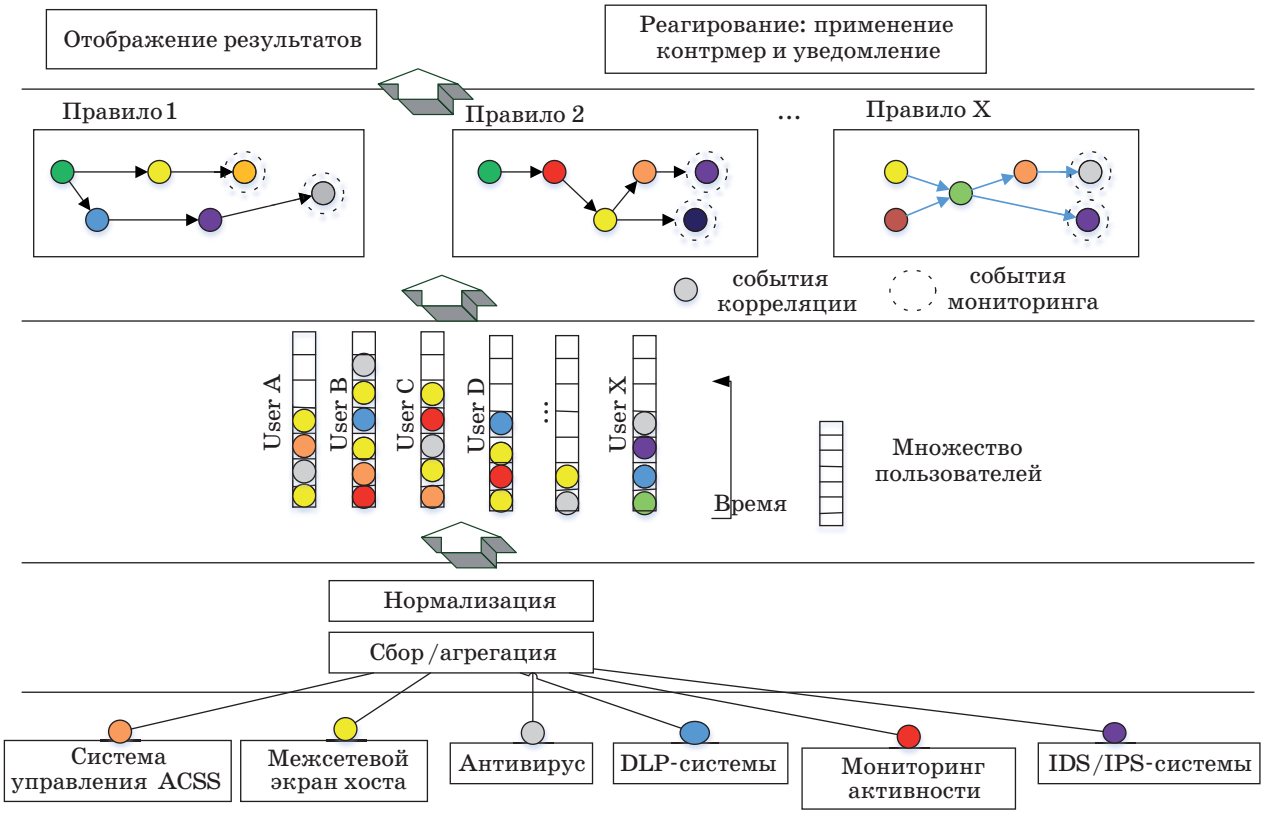
В настоящей работе рассматриваются основные алгоритмы корреляции событий, оценивается их способность выявлять сложные целенаправленные атаки, а также представляется существующее программное обеспечение, выполняющее корреляцию событий безопасности.

Алгоритмы на основе подобия событий безопасности

Алгоритмы на основе подобия событий безопасности базируются на вычислении показателя подобия, позволяющего выполнить сравнение двух событий безопасности или одного события с группой событий безопасности. Если уровень подобия превышает или равен некоторому заданному пороговому значению, то события объединяются в одно метасобытие. Таким образом, целью этих алгоритмов является агрегирование событий во времени или построение обобщающих иерархий событий. Основным преимуществом алгоритмов этой группы является отсутствие необходимости точного определения типа атак, и корреляция события может быть выполнена только на определении показателя подобия для атрибутов событий безопасности.

Определение подобия на основе правил

Основная идея этого подхода заключается в применении достаточно простых правил для описания взаимосвязей между атрибутами событий, которые могут быть связаны между собой. На рис. 1 представлена общая схема корреляции событий безопасности на основе правил [4].



■ ***Рис. 1.*** Общая схема корреляции событий безопасности на основе правил

■ ***Fig. 1.*** General scheme of rule-based event correlation

На первом этапе формируется множество правил, которые устанавливают связи между событиями безопасности для пользователей в системе, имеющей определенную роль в информационной системе. Далее при появлении контролируемых событий срабатывают определенные правила, которые в свою очередь могут инициировать срабатывание других правил, на основе которых делается заключение о выполнении атаки определенного типа. Конечный результат корреляции событий заключается в применении контрмер, заданных для каждого типа атаки, и его представлении в графическом виде администратору сетевой безопасности. Обычно описывают правила взаимосвязи между атрибутами данных трех уровней [5]:

— уровень данных (data level) — правила, работающие непосредственно с сырыми данными;

— уровень знаний (knowledge level) — правила, описывающие специфику предметной области и позволяющие работать с метасобытиями более высокого уровня;

— уровень управления (control level) — правила, непосредственно описывающие логический вывод на основе событий безопасности, т. е. это «сердце» модуля корреляции.

Очевидно, что эффективность таких систем зависит только от качества применяемых правил.

Определение подобия на основе кодовой книги

Клигер и др. [6] представили систему, целью которой является локализация возникающих проблем в системе на основе выбора некоторого подходящего подмножества событий-«симптомов», связанных с этими проблемами. Подмножество событий-«симптомов» и является содержанием кодовой книги. Для каждой проблемы создается некоторый двоичный вектор, который определяет, является некоторое событие признаком некоторой проблемы или нет, и записывается в кодовую книгу. Для выявления проблем все события, представленные в кодовой книге, отслеживаются в режиме реального времени. При наступлении некоторого события вектор события сравнивается с множеством векторов из кодовой книги, выбирается вектор, у которого расстояние Хемминга между ним и вектором события является минимальным. Благодаря такому решению система всегда выдает некоторое предположение о возможной проблеме. Очевидным недостатком данного подхода является невозможность учесть время между наступлением двух различных событий, что является важным параметром при установлении временных связей между событиями.

Определение подобия событий с использованием алгоритмов машинного обучения

В последнюю категорию вошли алгоритмы, в которых мера подобия между событиями определяется автоматически с помощью алгоритмов машинного обучения. В основном для решения этой задачи используются алгоритмы классификации из работы [7] и нейронные сети [8]. Так, например, для кластеризации событий используются деревья решения [7]. Причем данный алгоритм может применяться дважды — для определения похожих событий безопасности и для определения последствий атаки. Для корректного построения дерева решений и, соответственно, корректного функционирования самого алгоритма корреляции событий требуется обучающая выборка значительного размера, содержащая возможные сценарии атак. Появление новых данных, описывающих новые сценарии атак, требует переобучения модели анализа, что значительно снижает гибкость и расширяемость алгоритма.

В работе [8] классификация событий безопасности осуществляется с помощью нейронной сети и кластеризации, благодаря чему снижается число предупреждений, требующих ручной обработки. Однако авторы работ [1, 3] отмечают, что из-за отсутствия прозрачности в функционировании и обучении нейронных сетей они не очень популярны для построения инструментов корреляции событий.

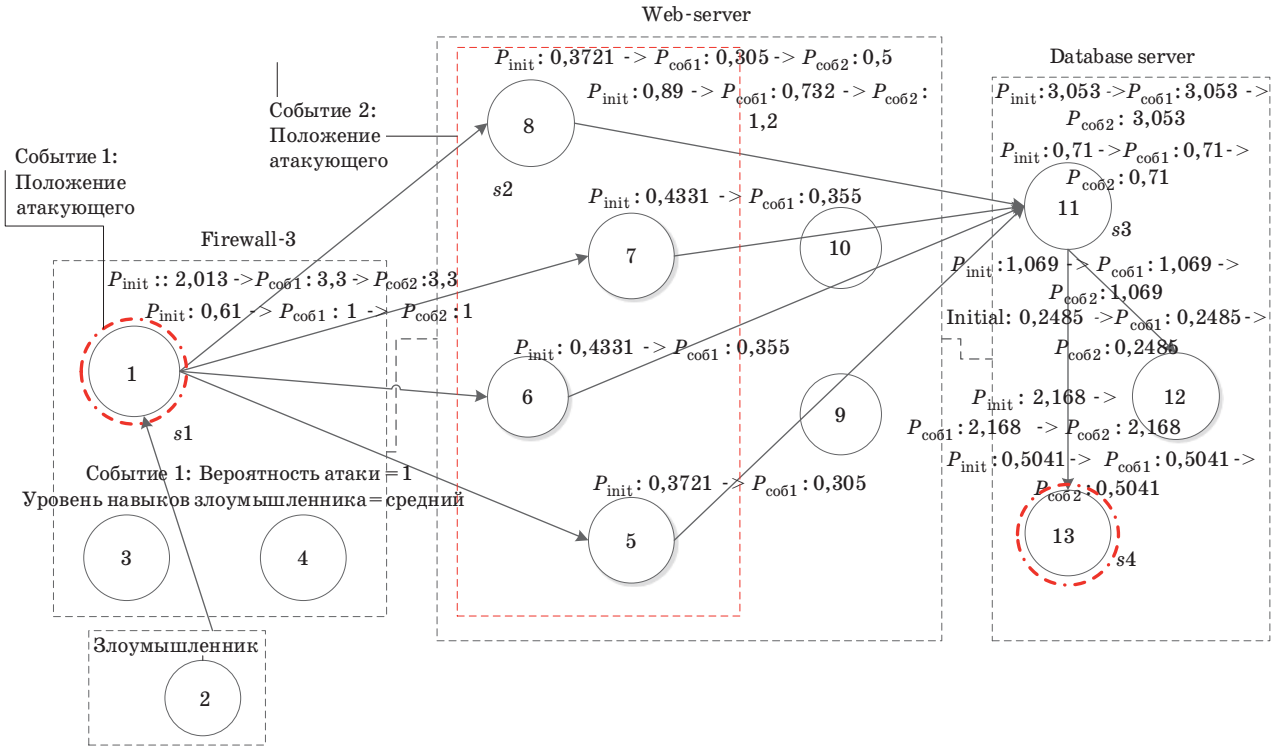
Алгоритмы на основе знаний

К этой категории относятся алгоритмы корреляции знаний, которые функционируют на основе описаний возможных сценариев атак или принципов функционирования контролируемой системы. Входящие в эту категорию алгоритмы можно разделить на две подгруппы.

Алгоритмы на базе предпосылок и последствий

Алгоритмы данной подкатегории отслеживают значение возникающих событий безопасности, оценивают состояние сети, после чего диагностируют наличие вторжения или какой-либо другой проблемы. Для того чтобы установить взаимосвязи между различными этапами атак, выражаемыми в виде цепочки некоторых событий безопасности, предполагается использование 1) базы данных, описывающих топологию сети, конфигурацию ее узлов; 2) базы знаний, которая для каждого события безопасности содержит описание его всех возможных предпосылок и последствий его наступления [9–11].

Обычно в этих алгоритмах события моделируются с помощью логики первого порядка.



■ ***Рис. 2.*** Пример сценария атаки
■ ***Fig. 2.*** Example of an attack scenario

Результат их работы может быть представлен в виде графа возможных событий и связей между ними. В работе [9] предложено расширение данного подхода для определения элементов атак, которые не были диагностированы датчиками безопасности. Сценарий атаки (рис. 2), построенный с помощью системы, представленной в работе [11], отображается в виде связного графа, вершинами которого являются атакующие действия злоумышленника, ребра обозначают возможную последовательность появления таких событий. Возможность выполнения атакующего действия связана с наличием уязвимостей программного обеспечения, сложность эксплуатации которых определяет исходную вероятность их наступления. Для каждого события исходная вероятность его наступления указана как P_{init}. Вершины графа ($s1$–$s4$) отражают последовательность действий злоумышленника, обладающего средним уровнем технических навыков. Вероятности $P_{соб1}$ и $P_{соб2}$ обозначают вероятности наступления событий после того, как в системе были зарегистрированы события 1 и 2, произошедшие на межсетевом экране Firewall-3 и на веб-сервере Web-server соответственно. Штрихпунктирной линией на графе выделены узлы, обозначающие исходное положение злоумышленника (вершина *1*) и наиболее вероятное конечное событие в результате атаки (вершина *13*).

Алгоритмы на основе сценариев атак

Основным назначением этого множества алгоритмов является определение многошаговых атак. Примеры таких алгоритмов представлены в работах [12–15]. Для описания сценариев атак предложены различные языки (Statl [14], ADeLe [15]), однако их общая идея заключается в описании этапа атаки, необходимых условий для его выполнения и цели. Алгоритмы этой группы оперируют знаниями более высокого уровня по сравнению с алгоритмами на основе предпосылок и последствий, поскольку последние функционируют на уровне событий безопасности.

Вероятностные алгоритмы корреляции событий безопасности

В основе этой группы алгоритмов лежит предположение, что атаки имеют общее статистическое распределение атрибутов, и корректная классификация атак может быть выполнена на основе оценки распределения значений атрибутов в сетевом трафике. В общем случае алгоритмы этой группы формируют базу причинно-следственных связей между различными событиями безопасности, анализируя частоту их возникновения в контролируемой системе во время периода обучения системы и строя возможные сценарии атаки. Непараметрические статистические алгоритмы не

требуют никакой априорной информации о возможных сценариях атак. Алгоритмы этой группы также могут быть разделены на две подгруппы. Алгоритмы первой подгруппы строят статистическую модель сетевого трафика, а алгоритмы второй подгруппы оценивают причинно-следственную связь между событиями безопасности.

Построение статистической модели сетевого трафика

Целью алгоритмов, представленных в работе [16], является создание статистической модели сетевого трафика, прогнозирование и исключение предсказуемых ситуаций. Важной особенностью данных алгоритмов является возможность диагностирования событий, которые возникают периодически по причине вероятных некорректных сетевых настроек или политик безопасности. Алгоритмы этой группы обычно не требуют данных о контролируемой системе, их функционирование определяется статистическими данными, сформированными на этапе обучения. В большинстве случаев обучение осуществляется в режиме реального времени, поэтому подстройка алгоритмов к изменениям в конфигурации информационной системы осуществляется достаточно просто и гибко.

В работе [17] представлен подход к корреляции событий безопасности, в основе которого лежат алгоритмы поиска ассоциативных правил для выявления событий безопасности, которые обычно возникают совместно. Важной особенностью данной подгруппы алгоритмов является определение приоритетов событий безопасности на основе того факта, является ли выявленная комбинация событий характерной для данной системы или представляет собой новый паттерн атаки. Кроме того, алгоритмы поиска ассоциативных правил могут быть использованы для формирования связанных метасобытий безопасности.

Оценка причинно-следственной связи между событиями

В основе алгоритмов этой подгруппы лежит построение возможной модели, определяющей корреляционные связи между событиями безопасности. Например, в работах [18–21] построение причинно-следственных связей между событиями осуществляется путем оценивания влияния заданного события в процессе предсказания появления других событий безопасности. Особенность предложенных методик заключается в том, что для построения моделей анализа не требуется дополнительная информация о контролируемой системе. Как и в предыдущей подгруппе алгоритмов, для получения точной и надежной модели требуется большое количество исторических данных со сценариями атак.

В работе [19] для построения вероятностных моделей используются байесовские сети.

Авторы работы [20] для определения вероятностных зависимостей между событиями безопасности используют скрытые марковские цепи. Отличительной характеристикой предложенного алгоритма является возможность оценки вероятности каждого сценария атак и выполнения каждого этапа атаки на основе предыдущих шагов. Вероятности оцениваются на основе обучающей выборки, например, исторических данных об атаках на контролируемую систему, поэтому для обучения модели анализа требуется большое количество данных, содержащих правильно категоризированные сценарии атак.

Сравнительный анализ рассмотренных подходов к корреляции событий безопасности

Чтобы описать достоинства и недостатки рассмотренных подходов для применения в системах управления информационной безопасностью, были выделены следующие критерии оценки:

— возможности, заключающиеся в способности алгоритмов агрегировать схожие события, выявлять последовательность событий от разных сенсоров безопасности и сетевых устройств, образующих единый сценарий атаки;

— необходимость применения базы знаний, определяющей корректность функционирования системы обнаружения атак;

— точность подхода, заключающаяся в способности алгоритмов обнаруживать атаки и прогнозировать их развитие;

— гибкость и расширяемость, оценивающая уровень адаптируемости модуля корреляции событий алгоритмов к появлению новых видов атак и возможность настройки параметров корреляции пользователем;

— вычислительная эффективность алгоритмов, определяющая мощность вычислительных ресурсов, требуемых для выполнения корреляции событий безопасности в целях выявления атак.

Алгоритмы на основе подобия событий безопасности в меньшей степени требуют контекстной информации о предметной области, о возможных сценариях атак в частности, поскольку они выполняют корреляцию данных на основе анализа подобия атрибутов события безопасности, что делает их более универсальными по сравнению с методиками корреляции на основе знаний, которые предполагают наличие данных о конфигурации устройств сети, ее топологии, установление зависимостей между используемыми сетевыми сервисами. Очевидно, результативность алгоритмов как на основе знаний, так

и на основе правил сильно зависит от корректности описания используемых правил, семантической нагрузки событий безопасности, поэтому их разработка требует непосредственного участия экспертов в информационной безопасности. Исключение составляют методики корреляции на основе правил, в которых взаимосвязи между атрибутами событий безопасности устанавливаются с помощью методик машинного обучения. Однако в этом случае результаты корреляции не являются «прозрачными» для конечного пользователя, и, как показали исследования различных SIEM-систем, механизмы валидации корректности функционирования моделей корреляции данных отсутствуют [22–25]. Определение существующих сценариев атак, как и установка предпосылок и последствий, является нетривиальной задачей, качество решения которой определяется в первую очередь полнотой исходных данных о предметной области. Исследования показали, что сложность графа атак, построенного для компьютерной сети, состоящей из n узлов, только на основе данных о топологии, конфигурации ее узлов без учета существующих зависимостей между сетевыми сервисами составляет $O(scn^2)$, где c — это среднее число уязвимостей для одного хоста, s — среднее число условий на хосте, обеспечивающих реализацию атаки [26]. Это определяет достаточно высокие требования к вычислительным мощностям устройств при выполнении корреляции событий в режиме реального времени. Вместе с тем именно алгоритмы на основе сценариев атак или графов атак способны выявлять сложные многошаговые атаки, объединяя множество событий безопасности, зарегистрированных на разных узлах компьютерной сети в различные периоды времени, в единую последовательность, описывающую действия злоумышленника.

Высокой точностью — низким уровнем ошибок, обнаружением ложноположительных срабатываний сенсоров безопасности — обладают алгоритмы, требующие исходных данных в виде экспертных знаний. В первую очередь к ним относятся алгоритмы на основе сценариев атак и предпосылок-последствий, во вторую — алгоритмы на основе подобия. Существенным недостатком этих двух подходов является их неспособность выявлять новые типы атак, использующие неизвестные на момент разработки модели корреляции данных уязвимости программного обеспечения, ошибки настроек сетевых устройств. Таким образом, для адаптации модуля корреляции к появлению новых, уже выявленных атак необходимо осуществлять регулярное обновление баз используемых правил, а в случае использования алгоритмов на основе сценариев атак и графов атак — обновлять структуру графа атак при любом изменении конфигурации компьютерной сети, настроек сетевых устройств и обнаруженных уязвимостей программного обеспечения.

Способностью обнаруживать новые сценарии атак обладают вероятностные методы корреляции, поскольку в их основе лежит построение статистической модели функционирования исследуемой сети, и любое отклонение от нее может быть расценено как возможное действие злоумышленника. Кроме того, они обладают достаточно высокой эффективностью [1]. А вот общая точность этой группы алгоритмов корреляции невысока в связи с изменчивостью модели «нормального» функционирования компьютерной сети.

Обобщенная сравнительная характеристика подходов дана в табл. 1.

Следует отметить, что на практике в основном применяются методы корреляции событий на основе правил подобия и кодовой книги [22–24,

■ ***Таблица 1***. Сравнительный анализ подходов к корреляции событий безопасности

■ ***Table 1***. Comparative analysis of approaches to the security events correlation

Характеристика	Алгоритмы		
	на основе подобия	на основе знаний	вероятностные
Комбинация событий безопасности от разных датчиков безопасности	Да	Да	Да
Требование предварительных знаний (обучение алгоритма)	Да	Да	Нет
Точность (обнаружение ложных событий безопасности)	Да	Да	Предположение
Обнаружение многошаговых атак	Едва ли	Да	Предположение
Обнаружение новых атак	Нет	Нет	Да
Уровень ошибок	Средний	Низкий	Высокий
Вычислительная эффективность	Высокая	Низкая	Средняя
Гибкость и расширяемость	Высокая	Высокая	Низкая

27], что объясняется в первую очередь их высокой вычислительной эффективностью и точностью, проблема обнаружения новых сценариев атак решается путем регулярного обновления баз знаний. Авторы считают достаточно перспективными также вероятностные методы корреляции, которые обладают как низкой вычислительной эффективностью, так и способностью к выявлению новых типов атак, что в сочетании с методиками на основе подобия позволит не допустить возникновения ситуаций пропуска атак в силу их новизны.

Программное обеспечение с открытым кодом, выполняющее корреляцию событий

Инструмент корреляции потоков событий Borealis

Инструмент корреляции событий Borealis [28] является распределенной системой корреляции потоковых данных в режиме реального времени. Одним из основных достоинств данной системы является оптимизация распределенной обработки, развертывание сети взаимодействующих модулей корреляции потоковых данных, которые осуществляют распределение входных данных по доступным вычислительным узлам, обеспечивая целостность данных даже при динамическом изменении сети модулей. Правила корреляции (запросы) описываются с помощью языка разметки XML.

Особенностью инструмента является возможность изменения запросов (правил) в динамике. Динамическое перестроение запросов необходимо при исправлении ошибок, возникающих в результате функционирования модуля корреляции, для получения более точных результатов. Кроме того, эта возможность полезна при обработке данных, которые могут поступать с определенными задержками, что достаточно характерно для датчиков, создающих потоковые данные.

Borealis имеет графическую подсистему, которая позволяет: 1) редактировать правила-запросы; 2) отслеживать топологию сети узлов Borealis в динамике.

Инструмент корреляции событий Simple Event Correlator

Инструмент Simple Event Correlator (SEC) [29] является движком корреляции события с открытым кодом, написанным на языке Perl. Использование языка программирования Perl обеспечивает кроссплатформенность приложения.

В основе корреляции событий лежит подход на основе правил. Правила могут быть описаны как в виде текста, так и с помощью конструкций языка Perl.

Инструмент SEC поддерживает достаточно большой набор базовых операций по корреляции событий: выявление события или пары событий по шаблону и реакция на него (SingleWithScript, SingleWithSuppress, Pair, PairWithWindow), отслеживание интенсивности появления события в течение некоторого интервала времени (SingleWithThreshold, SingleWith2Thresholds), запуск скриптов по графику и т. д.

Инструмент корреляции событий Esper

Инструмент Esper [30] является компонентом корреляции событий с открытым кодом, написанным на языке Java, и предназначен для разработки приложений, обрабатывающих события в режиме реального времени.

Esper является инструментом общего назначения и применяется для решения задач автоматизации и управления бизнес-процессами, обслуживания, мониторинга сетевого трафика и контроля приложений.

Правила создаются с помощью SQL-подобного языка, известного как EQL. Его конструкции обеспечивают обнаружение заданных шаблонов событий в режиме реального времени, поддерживают корреляцию событий во времени.

В табл. 2 представлены наиболее широко используемые инструменты общего назначения корреляции событий и их основные характеристики.

Все рассмотренные инструменты корреляции событий используют подход на основе правил.

■ ***Таблица 2***. Основные свойства инструментов корреляции

■ ***Table 2***. Basic features of correlation tools

Программное обеспечение	Корреляция событий на основе правил	Особенности
Borealis	На основе языка разметки XML	Распределенный процесс корреляции, изменение правил в динамике
SEC	В текстовом виде	Сложная корреляция данных, представленных в текстовом виде, на основе простых операций
Esper	Sql-подобный запрос	Высокопроизводительная обработка потоковых данных (событий)

Выделенные особенности могут быть полезны при выборе инструмента для решения конкретных задач.

Заключение

Представлены алгоритмы событий безопасности, проведен сравнительный анализ подходов к корреляции событий. Рассмотрено существующее программное обеспечение, осуществляющее корреляцию событий безопасности, проведен сравнительный анализ инструментов корреляции и их характеристик. Результаты исследования будут использованы при разработке механизмов защиты облачных вычислительных сред.

Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации в рамках государственного задания «Организация научных исследований», задание #2.6113.2017/6.7, а также гранта РФФИ № 16-07-00625.

Литература

1. **Mirheidari S. A., Arshad S., Jalili R.** Alert Correlation Algorithms: A Survey and Taxonomy // Proc. of 5th Intern. Symp. on Cyberspace Safety and Security (CSS 2013), Zhangjiajie, China, Nov. 13–15, 2013. LNCS, 2013. Vol. 8300. P. 183–197.
2. **Valdes A., Skinner K.** An Approach to Sensor Correlation // Proc. of the Recent Advances in Intrusion Detection (RAID-2000), Touluse, 2000. https://www.researchgate.net/profile/Alfonso_Valdes/publication/228523518_An_approach_to_sensor_correlation/links/56d4437d08ae868628b24ba8.pdf (дата обращения: 27.06.2017).
3. **Mueller A.** Event Correlation Engine. https://www.open.ch/_pdf/internships/EventCorrelationEngine_AndreasMueller.pdf (дата обращения: 27.06.2017).
4. **Kang D., Na J.** A Rule Based Event Correlation Approach for Physical and Logical Security Convergence // IJCSNS Intern. Journal of Computer Science and Network Security. 2012. Vol. 12. N 1. P. 28–32.
5. **Elshoush H. T., Osman I. M.** Intrusion Alert Correlation Framework: An Innovative Approach // IAENG Transactions on Engineering Technologies. 2013. Vol. 229. P. 405–420.
6. **Klinger S., Yemini S., Yemini Y., Ohsie D., Stolfo S.** A Coding Approach to Event Correlation // Proc. of the Fourth Intern. Symp. on Integrated Network Management IV/ Adarshpal S. Sethi, Yves Raynaud, and Fabienne Faure-Vincent (Eds.). London: Chapman & Hall, 1995. P. 266–277.
7. **Dwivedi N., Tripathi A.** Event Correlation for Intrusion Detection Systems // Computational Intelligence & Communication Technology (CICT): 2015 IEEE Intern. Conf. IEEE, 2015. P. 133–139.
8. **Kidmose E., Stevanovic M., Pedersen J. M.** Correlating Intrusion Detection Alerts on Bot Malware Infections using Neural Network //Cyber Security and Protection of Digital Services (Cyber Security): 2016 Intern. Conf. IEEE, 2016. P. 1–8.
9. **Xuewei F., et al.** An Approach of Discovering Causal Knowledge for Alert Correlating Based on Data Mining //Dependable, Autonomic and Secure Computing (DASC): 2014 IEEE 12th Intern. Conf. IEEE, 2014. P. 57–62.
10. **Wang C., Chiou Y.** Alert Correlation System with Automatic Extraction of Attack Strategies by Using Dynamic Feature Weights // Intern. Journal of Computer and Communication Engineering. 2016. Vol. 5. N 1. P. 1–10.
11. **Kotenko I., Doynikova E.** Evaluation of Computer Network Security based on Attack Graphs and Security Event Processing // Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications. 2014. Vol. 5. N 3. P. 14–29.
12. **Katz G., Elovici Y., Shapira B.** CoBAn: A Context Based Model for Data Leakage Prevention // Information Sciences. 2014. Vol. 262. P. 137–158.
13. **Morin B., Me L., Debar H., Ducasse M.** A Logic-Based Model to Support Alert Correlation in Intrusion Detection // Information Fusion. 2009. Vol. 10. N 4. P. 285–299.
14. **Eckmann S. T., Vigna G., Kemmerer R. A.** An Attack Language for State-Based Intrusion Detection // Journal of Computer Security. 2002. N 10 (1–2). P. 71.
15. **Totel E., Vivinis B.** A Language Driven Intrusion Detection System for Event and Alert Correlation // Security and Protection in Information Processing Systems. 2004. Vol. 147. P. 208–224.
16. **Viinikka J., Debar H., Me L., Lehikoinen A., Tarvainen M.** Processing Intrusion Detection Alert Aggregates with Time Series Modelling // Information Fusion. 2009. Vol. 10. N 4. P. 312–324.
17. **Treinen J., Thurimella R.** A Framework for the Application of Association Rule Mining in Large Intrusion Detection Infrastructures // Recent Advances in Intrusion Detection (RAID). 2006. Vol. 4219. P. 1–18.
18. **Jakobson G.** Mission Resilience // Cyber Defense and Situational Awareness. — Springer International Publishing, 2014. — P. 297–322.
19. **Gao J., Jiang G., Chen H., and Han J.** Modeling Probabilistic Measurement Correlations for Problem Determination in Large-Scale Distributed Systems // Proc. IEEE, Montreal, Canada, June 22–26, 2009. P. 623–630.
20. **Naukudkar K. B., Ambawade D. D., Bakal J. W.** Enhancing Performance of Security Log Analysis using Correlation-Prediction Technique // Proc. of Intern. Conf. on ICT for Sustainable Development, Singapore, Feb. 26, 2016. Springer, 2016. Vol. 409. P. 635–643.

21. **Schutte J., Rieke R., Winkelvos T.** Model-based Security Event Management // Proc. of the 6th Intern. Conf. on Mathematical Methods, Models and Architectures for Computer Network Security: Computer Network Security (MMM-ACNS'12). Springer-Verlag, Berlin, Heidelberg, 2012. P. 181–190.
22. HPE ArcSight ESM. https://saas.hpe.com/ru-ru/software/siem-security-information-event-management (дата обращения: 25.06.2017).
23. Splunk Enterprise Security. https://www.splunk.com/en_us/products/premium-solutions/splunk-enterprise-security.html (дата обращения: 25.06.2017).
24. IBM QRadar SIEM. http://www-03.ibm.com/software/products/ru/qradar-siem (дата обращения: 25.06.2017).
25. **Walton S., Maguire E., Chen M.** A Visual Analytics Loop for Supporting Model Development //Proc. of 2015 IEEE Symp. on Visualization for Cyber Security (VizSec), Chicago, IL, 2015. P. 1–8.
26. **Котенко И. В., Новикова Е. С.** Визуальный анализ защищенности компьютерных сетей // Информационно-управляющие системы. 2013. № 3. С. 56–61.
27. **Шелестова О.** Корреляция SIEM — это просто. Сигнатурные методы. http://www.securitylab.ru/analytics/431459.php (дата обращения: 26.06.2017).
28. Borealis Distributed Stream Processing Engine. http://cs.brown.edu/research/borealis/public/ (дата обращения: 10.12.2016).
29. SEC — simple event correlator. https://simple-evcorr.github.io/ (дата обращения: 10.12.2016).
30. Esper: Event Processing for Java. http://www.espertech.com/products/esper.php (дата обращения: 10.12.2016).

UDC 004.056
doi:10.15217/issn1684-8853.2017.5.95

A Survey of Security Event Correlation Techniques for Cloud Computing Environment Security

Novikova E. S.[a,b], PhD, Tech., Assosiate Professor, novikova.evgenia123@gmail.com
Bekeneva Ya. A.[a], Engineer, yana.barc@mail.ru
Shorov A. V.[a], PhD, Tech., Leading Researcher, ashxz@mail.ru
Fedotov E. S.[a], Post-Graduate Student, fedotov_e1290@mail.ru
[a]Saint-Petersburg State Electrotechnical University «LETI», 5, Prof. Popov St., 197376, Saint-Petersburg, Russian Federation
[b]Saint-Petersburg Institute for Informatics and Automation of the RAS, 39, 14 Line, V. O., 199178, Saint-Petersburg, Russian Federation

Introduction: The increased complexity of attacks on cloud infrastructures leads to the necessity to develop security mechanisms able to analyze security events, including those separated in time. In addition, there is a need to verify the reliability of generated security events, as well as to assess the criticality of a security event in the context of the monitored asset criticality. This defines a need for a security event correlation module as a core element of the system. **Purpose:** We analyze the developed event correlation techniques and assess their applicability in cloud infrastructure. **Results:** The analysis of the major event correlation algorithms and available software has shown that there are three main approaches to the development of such algorithms: the approach on the base of security event similarity, the knowledge-based approach and probabilistic approaches. The following assessment criteria for the comparison of the existing approaches have been defined: the ability to correlate events from heterogeneous data sources, the requirements to prior knowledge base, the event correlation accuracy, and the ability to detect novel and multistep attacks. The results of the comparison analysis are presented. **Practical relevance:** The results of the research can be used in the development of protection mechanisms against targeted persistent attacks, securing cloud computing environment. The usage of security event correlation techniques enables security tools to prioritize security events more accurately and respond timely.

Keywords — Cloud Technology Security, Security Events, Correlation of Security Events, Event Correlation Algorithms.

References

1. Mirheidari S. A., Arshad S., Jalili R. Alert Correlation Algorithms: A Survey and Taxonomy. *Proc of 5th Intern. Symp. on Cyberspace Safety and Security (CSS 2013)*, Zhangjiajie, China, November 13–15, 2013, LNCS, 2013, vol. 8300, pp. 183–197.
2. Valdes A., Skinner K. An Approach to Sensor Correlation *Proc. of the Recent Advances in Intrusion Detection (RAID-2000)*, Touluse, 2000. Available at: https://www.researchgate.net/profile/Alfonso_Valdes/publication/228523518_An_approach_to_sensor_correlation/links/56d-4437d08ae868628b24ba8.pdf (accessed 27 June 2017).
3. Mueller A. *Event Correlation Engine*. Available at: https://www.open.ch/_pdf/internships/EventCorrelationEngine_AndreasMueller.pdf (accessed 27 June 2017).
4. Kang D., Na J. A Rule Based Event Correlation Approach for Physical and Logical Security Convergence. *IJCSNS Intern. Journal of Computer Science and Network Security*, 2012, vol. 12, no. 1, pp. 28–32.
5. Elshoush H. T., Osman I. M. Intrusion Alert Correlation Framework: An Innovative Approach. *IAENG Transactions on Engineering Technologies*, 2013, vol. 229, pp. 405–420.
6. Klinger S., Yemini S., Yemini Y., Ohsie D., Stolfo S. A Coding Approach to Event Correlation. *Proc. of the Fourth Intern. Symp. on Integrated Network Management IV*, Adarshpal S. Sethi, Yves Raynaud, and Fabienne Faure-Vincent (Eds.), 1995, pp. 266–277.
7. Dwivedi N., Tripathi A. Event Correlation for Intrusion Detection Systems. *IEEE Intern. Conf. "Computational Intelligence & Communication Technology" (CICT)*, 2015, pp. 133–139.
8. Kidmose E., Stevanovic M., Pedersen J. M. Correlating Intrusion Detection Alerts on Bot Malware Infections using Neural Network. *Intern. Conf. "Cyber Security and Protection of Digital Services (Cyber Security)"*, 2016, pp. 1–8.
9. Xuewei F., et al. An Approach of Discovering Causal Knowledge for Alert Correlating Based on Data Mining. *IEEE*

12th Intern. Conf. "Dependable, Autonomic and Secure Computing (DASC)", 2014, pp. 57–62.
10. Wang C., Chiou Y. Alert Correlation System with Automatic Extraction of Attack Strategies by Using Dynamic Feature Weights. *Intern. Journal of Computer and Communication Engineering*, 2016, vol. 5, no. 1, pp. 1–10.
11. Kotenko I., Doynikova E. Evaluation of Computer Network Security based on Attack Graphs and Security Event Processing. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, 2014, vol. 5, no. 3, pp. 14–29.
12. Katz G., Elovici Y., Shapira B. CoBAn: A Context Based Model for Data Leakage Prevention. *Information Sciences*, 2014, vol. 262, pp. 137–158.
13. Morin B., Me L., Debar H., Ducassé M. A Logic-Based Model to Support Alert Correlation in Intrusion Detection. *Information Fusion*, 2009, vol. 10, no. 4, pp. 285–299.
14. Eckmann S. T., Vigna G., Kemmerer R. A. An Attack Language for State-Based Intrusion Detection. *Journal of Computer Security*, 2002, no. 10(1–2), p. 71.
15. Totel E., Vivinis B. A Language Driven Intrusion Detection System for Event and Alert Correlation. *Security and Protection in Information Processing Systems*, 2004, vol. 147, pp. 208–224.
16. Viinikka J., Debar H., Me L., Lehikoinen A., Tarvainen M. Processing Intrusion Detection Alert Aggregates with Time Series Modelling. *Information Fusion*, 2009, vol. 10, no. 4, pp. 312–324.
17. Treinen J., Thurimella R. A Framework for the Application of Association Rule Mining in Large Intrusion Detection Infrastructures. *Recent Advances in Intrusion Detection (RAID)*, 2006, vol. 4219, pp. 1–18.
18. Jakobson G. Mission Resilience. In: *Cyber Defense and Situational Awareness*. Springer International Publishing, 2014, pp. 297–322.
19. Gao J., Jiang G., Chen H., Han J. Modeling Probabilistic Measurement Correlations for Problem Determination in Large-Scale Distributed Systems. *Proc. IEEE*, Montreal, Canada, June 22–26, 2009, IEEE, 2009, pp. 623–630.
20. Naukudkar K. B., Ambawade D. D., Bakal J. W. Enhancing Performance of Security Log Analysis Using Correlation-Prediction Technique. *Proc. of Intern. Conf. on ICT for Sustainable Development*, Singapore, February 26, 2016, Springer, 2016, vol. 409, pp. 635–643.
21. Schutte J., Rieke R., Winkelvos T. Model-based Security Event Management. *Proc. of the 6th Intern. Conf. on Mathematical Methods, Models and Architectures for Computer Network Security: Computer Network Security (MMM-ACNS'12)*, Springer-Verlag, Berlin, Heidelberg, 2012, pp. 181–190.
22. *HPE ArcSight ESM*. Available at: https://saas.hpe.com/ru-ru/software/siem-security-information-event-management (accessed 25 June 2017).
23. *Splunk Enterprise Security*. Available at: https://www.splunk.com/en_us/products/premium-solutions/splunk-enterprise-security.html (accessed 25 June 2017).
24. *IBM QRadar SIEM*. Available at: http://www-03.ibm.com/software/products/ru/qradar-siem (accessed 25 June 2017).
25. Walton S., Maguire E., Chen M. A Visual Analytics Loop for Supporting Model Development. *Proc. of 2015 IEEE Symp. on Visualization for Cyber Security (VizSec)*, Chicago, IL, 2015, pp. 1–8.
26. Kotenko I. V., Novikova E. S. The Visual Analysis For Computer Network Security Assessment. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2013, no. 3, pp. 56–61 (In Russian).
27. Shelestova O. *Korreliatsiia SIEM — eto prosto. Signaturnye metody* [SIEM Correlation it's Easy. Signature Methods]. http://www.securitylab.ru/analytics/431459.php (accessed 26 June 2017).
28. *Borealis Distributed Stream Processing Engine*. Available at: http://cs.brown.edu/research/borealis/public/ (accessed 26 June 2017).
29. *SEC — Simple Event Correlator*. Available at: https://simple-evcorr.github.io/ (accessed 26 June 2017).
30. *Esper: Event Processing for Java*. Available at: http://www.espertech.com/products/esper.php (accessed 26 June 2017).