УДК 004.056.53 **DOI** 10.15622/sp.49.11

А.В. Федорченко, Д.С. Левшун А.А. Чечулин, И.В. Котенко

АНАЛИЗ МЕТОДОВ КОРРЕЛЯЦИИ СОБЫТИЙ БЕЗОПАСНОСТИ В SIEM-СИСТЕМАХ. ЧАСТЬ 2

Федорченко А.В., Левшун Д.С., Чечулин А.А., Котенко И.В. **Анализ методов корреляции событий безопасности в SIEM-системах. Часть 2.**

Аннотация. Статья является продолжением описания исследований, посвященных анализу методов корреляции событий безопасности в системах управления информацией и событиями безопасности (SIEM-системах). В данной части рассматриваются методы непосредственной корреляции событий безопасности, применяемых на этапах, описанных в предыдущей статье. Приводится классификация рассматриваемых методов корреляции и результаты анализа их достоинств и недостатков, а также оценивается эффективность их применения на различных этапах процесса корреляции.

Ключевые слова: методы корреляции данных, события безопасности, анализ событий безопасности, системы оценки защищенности, SIEM-системы.

1. Введение. В первой части описания проведенных исследований корреляция событий безопасности была рассмотрена как один из важнейших процессов, выполняемых в системе управления информацией и событиями безопасности (Security Information and Event Management, SIEM) [1, 2]. Данный класс систем является перспективным и востребованным на мировом рынке решений, обеспечивающих мониторинг защищенности компьютерных инфраструктур.

Стоит кратко упомянуть о целях и задачах проведенного исследования, а также о результатах, описанных в предыдущей статье. Так как модуль корреляции является неотъемлемой частью любой SIEM-системы, а возложенные на него функции влияют на результат работы всей системы в целом, основная цель работы заключается в изучении корреляции как составного процесса, так и рассмотрении основных применяемых математических методов и технических подходов. Задача детального анализа конкретных методов корреляции необходима для выявления их достоинств и недостатков, что позволит определить эффективность применения каждого из них на различных этапах процесса корреляции, определенных в первой части. Были выделены следующие этапы: (1) нормализация; (2) предобработка; (3) анонимизация; (4) агрегация и фильтрация; (5) восстановление хода атаки; (6) восстановление сессии атаки; (7) определение источника и цели атаки; (8) многошаговая корреляция; (9) анализ ущерба; (10) приоритезация; (11) фильтрация на основе ранжирования. Представленные этапы были разделены на 5 уровней, определяющих

ход обработки событий безопасности, однако данная схема не исключает обратных и цикличных маршрутов следования событий между уровнями. Вместе с тем наряду с общей схемой процесса корреляции большую значимость при обработке событий безопасности имеют используемые методы поиска взаимосвязей и их параметров над множеством входных данных (событий и информации безопасности).

Статья имеет следующую структуру. В разделе 2 описываются релевантные работы, посвященные обзору методов корреляции, их классификации и практического применения. На основе данных статей были получены основные сведения о методах корреляции. Раздел 3 посвящен рассмотрению методов корреляции, применяемых на разных этапах процесса обработки событий безопасности. Выполняется классификация методов, описываются достоинства, недостатки и способы применения каждого из них. Также дается общее представление способов обучения системы корреляции, основанной на приведенных методах.

2. Релевантные работы. Классификация методов корреляции событий безопасности является важным аспектом исследуемой предметной области, поскольку на основе ее можно делать заключение о возможности применения как отдельно взятого метода, так и их комбинирования в модуле корреляции SIEM-системы. По тематике изучения методов корреляции написано достаточное количество работ для их исследования. Ниже представлены работы, наиболее полно описывающие различные методы корреляции для получения ясной картины их функционирования и оценки эффективности применения на различных этапах процесса корреляции.

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

В [4] также приводится собственная классификация систем корреляции, согласно которой выделяются следующие типы систем*:* (1) *системы, основанные на правилах* (*rule-based*); (2) *системы на базе*

кодовых книг (*codebook*) и (3) системы *с использованием интеллектуальных методов* (*artificial intelligence*). Однако, в рамках представленной работы, особенности и недостатки каждой из перечисленных типов систем не приводятся.

В [5] описываются две модели корреляции событий безопасности: (1) *причинно-следственная* и (2) *временная*. Данные модели основаны на определении соответствия между событиями и упорядочивании событий в хронологическом порядке соответственно. В работе считается, что в системах корреляции обычно производится сопоставление указанных моделей с топологией анализируемой сети. Вместе с этим в статье приводится доказательство использования *причинно-следственной* модели корреляции в составе *временной* модели. Стоит отметить, что простота приведенных моделей процесса корреляции событий безопасности делает их неприменимыми в качестве механизма непосредственного описания взаимосвязей между событиями в реальных системах, а сама работа нацелена скорее *не на пользователей*, а на *разработчиков* систем корреляции.

Работа [6] посвящена разработке сервис-ориентированного приложения корреляции событий и содержит описание основных применяемых методов. Особенностью данной работы является описание примеров реализации методов корреляции как на этапе проектирования, так и в ходе непосредственного функционирования разработанного прототипа. Выделяются следующие методы корреляции: метод на основе моделирования (model-based reasoning, MBR); правило-ориентированный метод (rule-based reasoning, RBR); метод на основе кодовой книги (codebook); метод рассуждений на основе прецедентов (case-based reasoning, CBR); метод активного исследования (active probing). В работе перечисленные методы корреляции были сравнены по различным свойствам, определена актуальность использования каждого метода в автоматизированном сервисе корреляции. Например, свойство возможности сопровождения метода (*maintenance*) характеризует способность модификации метода. Свойство поддержки моделирования (*modeling*) отражает возможность моделирования метода. Под свойством надежности (*robustness*) понимается отказоустойчивость встроенных механизмов исправления ошибок. Свойство производительности (*performance*) отражает быстродействие метода на основе сложности используемых им алгоритмов. По выделенным свойствам наиболее выигрышным является метод рассуждений на основе прецедентов. Однако в предлагаемой архитектуре сервиса корреляции данный метод используется только при ошибках в результате работы ранее

используемых методов (правило-ориентированном и на основе моделирования). Также отмечается, что ни один из методов не годится для исключительного (единственного) использования в разрабатываемом сервисе корреляции.

В диссертации [7] приведено описание модельно-методического аппарата для корреляции событий безопасности. Работа включает разбор этапов процесса корреляции, применяемых методов и типов операций, производимых над данными, различных систем корреляции событий безопасности и их сравнительный анализ, а также форматы описания событий. Важным элементом данной диссертации является оценка преимуществ и недостатков использования различных функциональностей и режимов работы системы корреляции, таких как самообучение и использование внешних знаний, реальное время и сохранение данных (с сохранением состояния и без), активный и пассивный режимы работы, централизованное и децентрализованное управление, глубокий анализ и поверхностное исследование. Несмотря на то что данные пары функциональностей и режимов работы обладают противоположными качествами, в определенных случаях целесообразно, чтобы система обладала обоими свойствами из одной пары. Данная работа также достаточно полно раскрывает детали процесса корреляции с точки зрения применяемых методов и подходов. Выделяются следующие методы корреляции: (1) *машина конечных состояний*; (2) *правила*; (3) *поведенческий анализ*; (4) *моделирование*; (5) *кодовая книга*; (6) *голосование*; (7) *явное обозначение ошибки*; (8) *граф зависимостей*; (9) *байесовская сеть*; (10) *нейронная сеть*. Данные методы по своей сущности имеют одно или несколько формальных оснований, таких как теория графов, четкая и нечеткая логика, теория вероятности, математическая статистика, машинное обучение и интеллектуальный анализ данных.

В [8] предлагается различать методы корреляции событий безопасности по следующим ортогональным критериям: (1) *способ корреляции событий безопасности*; (2) *уровень корреляции событий безопасности*; (3) *используемые форматы данных*. Авторы работы отмечают, что форматы данных получаемых пакетов, потоков и событий безопасности должны быть определяемыми. Приводится классификация способов непосредственной корреляции, которая различает *сигнатурные* и *бессигнатурные* (на основе обнаружения аномалий) алгоритмы. При этом в категории бессигнатурных алгоритмов выделяются (1) подходы, основанные на *спецификации* (specification-based), выполняющиеся, как правило, в ручном или полуавтоматическом режиме, и (2) подходы,

базирующиеся на *интеллектуальном анализе данных* (data-mining-based). В работе отмечается, что бессигнатурные алгоритмы могут выбираться в зависимости от анализируемого трафика. Если *нормальное поведение системы* используется в качестве *данных обучения*, то генерация событий безопасности будет соответствовать *несовпадению* наблюдаемых данных с шаблоном обучения. Если же *аномальная активность системы* используется в качестве *данных обучения*, то генерация событий безопасности будет соответствовать *совпадению* наблюдаемых данных с шаблоном обучения.

На основе данного анализа работ была разработана собственная классификация методов корреляции, и приведено описание и способы применения наиболее употребляемых из них. Полученные результаты приводятся в разделе 3.

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

В рамках процесса корреляции, от получения разноуровневых событий до формирования результатов, задачами методов корреляции являются [7-10]:

– преобразование данных для их понимания системой анализа защищенности компьютерной сети или SIEM-системой;

– преобразование данных от уровня к уровню для определения общего состояния анализируемой инфраструктуры;

– автоматизированное определение наиболее важных связей событий для их дальнейшего применения при анализе состояния инфраструктуры (самостоятельное обучение системы);

– приведение данных к виду, понятному администратору безопасности;

– сокращение данных до объема, приемлемого для обработки администратором безопасности.

Сами методы, применяемые в процессе корреляции, можно классифицировать по следующим параметрам:

(1) по возможности изменения способа обработки корреляционных признаков:

– статические;

– динамические.

(2) по способу изменения корреляционных признаков:

– самостоятельно изменяемые (самообучаемые);

– изменяемые вручную;

– неизменяемые (необучаемые, фиксированные на начальном этапе).

(3) по типу вычисления результата:

– упорядоченные;

– вероятностные;

– смешанные.

(4) по возможности определения пути вычисления результата:

– определяемые;

– неопределяемые.

(5) по варианту оперирования корреляционными признаками:

– использующие;

– определяющие;

– совместные.

Параметр возможности изменения способа обработки корреляционных признаков отражает способность метода использовать операции различного характера для обработки корреляционных признаков. В данном случае при математическом описании методов корреляции следует обратить внимание на результирующую функцию каждого метода. Описываемая характеристика является показателем возможности изменения способа обработки аргументов данной функцией. Так, например, в методе на основе правил обработка аргументов функции будет заключаться в сравнении их значений со значениями, заключенными в условиях. Таким образом, метод не подразумевает нескольких возможных способов обработки входных данных и поэтому является статичным по возможности изменения корреляционных признаков. В свою очередь, нейронные сети, имеющие сумматорную и активационную функции, которые в конечно счете и являются результирующими, могут реализовывать различные способы обработки корреляционных признаков.

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

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

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

Возможность определения пути вычисления результата отражает способность обратимости алгоритмов, реализующих метод. Например, в случае метода на основе конечного автомата и правило-ориентированного метода, при известных входных и выходных

параметрах путь выполнения алгоритма будет всегда однозначным и обратимо вычисляемым, то есть вычисляемым как от входных до выходных данных, так и наоборот. В случае с самообучаемыми методами корреляции обратное преобразование от выходных к входным данным является трудновыполнимым или невыполнимым вовсе, так как входные данные уже могли повлиять на чувствительные элементы алгоритма (например, веса), а сохранение предыдущих состояний данных элементов не имеет особого смысла. Данная характеристика является полезной для определяемых методов при отладке и тестировании их конкретных реализаций, так как позволяет локализовать ошибку в работе алгоритма. Для неопределяемых методов поиск ошибки обратным преобразованием выходных данных невыполним.

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

В ходе анализа работ, описывающих архитектуры, алгоритмы, методы и системы корреляции событий и предупреждений безопасности, было выделено пять методов, являющихся наиболее используемыми в области SIEM-систем, систем обнаружения и предотвращения вторжений и др. Данные методы отличаются друг от друга по ряду характеристик, учтенных в описанных признаках классификации. Также данная разновидность позволяет использовать представленные методы в разных этапах процесса корреляции с учетом конкретных решаемых задач.

В таблице 1 представлены выделенные методы корреляции событий безопасности и их классификация в соответствии с рассмотренными признаками.

Среди представленных в классификации методов не были добавлены такие методы, которые только определяют корреляционные признаки. Данное положение обусловлено ограничением применения такого метода только в рамках обучения системы, что выходит за рамки данной статьи. К таким методам относятся методы кластеризации, построения деревьев решений, классификации, алгоритмы которых позволяют производить оценку качества выделенных корреляционных признаков, изначальное задание глубины анализа и многие другие характеристики.

Таблица 1. Классификация методов корреляции

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

Методы, являющиеся как использующими, так и определяющими (то есть совместными) по оперированию с корреляционными признаками, применяются как на этапах обработки данных процесса корреляции (применения признаков обработки), так и на этапах обучения системы корреляции (определения признаков обработки).

Предложены также такие методы корреляции, как кодовая книга [4, 6], на основе сценариев/шаблонов атак [4, 11], на основе модели состояния [6, 7], голосование [7], явное изолирование ошибок [7], на основе графов зависимостей [7, 12], генетические алгоритмы [7, 10], школьная доска [7], на основе контекстно-независимой грамматики [7], на основе стандартного и аномального поведения [4, 11], на основе иммунных систем [10], временно-ориентированные [11], на основе нечеткой логики [7, 13, 14], на основе схожести и др. [4, 6, 7, 10, 11, 15].

Также используется ряд методов интеллектуального анализа событий для выявления корреляционных признаков. Стоит отметить, что несмотря на большое количество существующих методов корреляции, ряд методов могут быть логически преобразованы в другие методы. Например, метод на основе графа зависимостей, метод конечных состояний и некоторые другие методы можно выразить правило-ориентированном методом.

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

4.1. Метод на основе машины конечных состояний (конечный автомат). Данный метод основан на построении модели графа переходов между состояниями анализируемой инфраструктуры. В качестве условий перехода выступают определенные параметры событий, тогда как само состояние определяет операции над анализируемым потоком событий. Метод конечных состояний включает в себя [7, 10]: (1) множество возможных входных событий (входной алфавит); (2) множество возможных выходных событий (выходной алфавит); (3) множество возможных состояний; (4) начальное состояние и (5) функцию перехода между состояниями.

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

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

Описанный метод наиболее подходит для идентификации «вредных» (опасных, предупреждающих) состояний системы при мониторинге анализируемой инфраструктуры [7]. В рамках общего процесса корреляции реализация данного метода возможна на этапах агрегации и определения источника и цели атаки за счет предопределенных состояний. На этапах, использующих более

сложные корреляционные признаки, метод конечных состояний применять затруднительно.

4.2. Правило-ориентированный метод. Данный метод является классическим и широко-распространенным не только в SIEM-системах, но и системах обнаружения и предотвращения вторжений, межсетевых экранах и антивирусных решениях. В основе данного метода лежат правила, имеющие понятный системе синтаксис и семантику [4, 6, 7, 11, 16]. Правило в данном методе является самостоятельной оперативной единицей (то есть операция может осуществиться за счет лишь одного правила). Каждое правило состоит из условия, проверяемого для входных данных по корреляционным признакам, и действия над поступившими данными в случае выполнения условия. Все правила для каждой операции обработки данных находятся в хранилище правил. При поступлении данных на вход конкретной операции, они проходят проверку на предусловие, а именно, соответствие корреляционных признаков входной информации по отношению к правилам из хранилища. Также правила можно разделять на простые и сложные. Например, простым правилом можно назвать такую запись в таблице, для положительного исхода которой (применения правила) достаточно выполнения одного предусловия. В свою очередь, к сложным правилам относится набор из простых и сложных правил, связанных логическими операторами (И, ИЛИ, НЕ) и их комбинациями. Правила в хранилище можно добавлять, удалять и изменять в процессе работы всей системы. Однако изменение правил с помощью самих правил не предусмотрено, поэтому данный метод является статическим и изменяемым вручную. Последнее свойство также обусловлено сложностью составления самих правил. Правило-ориентированный метод является упорядоченным, определяемым и использующим корреляционные признаки, ввиду использования четкой логики выполнения правил, конечности их множества и отсутствия возможности применения для анализа данных с целью определения корреляционных признаков.

Правило-ориентированный метод применим в процессе корреляции на этапах нормализации, анонимизации, агрегации и фильтрации. Преимуществом описанного метода является строгое соблюдение условий при принятии решений, однако это не исключает логических ошибок, связанных с пересечением правил, например, при удовлетворении предусловий сразу нескольких правил, с противоречивыми результатами их выполнения. Недостатком данного метода является большая емкость правил и сложность их составления для наиболее рациональной обработки как с точки зрения достижения

оптимальной точности анализа, так и минимизации затрачиваемых ресурсов на его выполнение.

4.3. Метод рассуждений на основе прецедентов. В основу данного метода положена ситуационная модель, характеризующая поведение анализируемой инфраструктуры. Данная модель строится по обучающему множеству ситуаций (прецедентов), на основе которой определяется характер текущего поведения анализируемой инфраструктуры на этапе работы системы корреляции. Построение модели основано на использовании принципа адаптации. Данный принцип заключается в наполнении хранилища прецедентов (случаев) записями с возможными решениями. При поступлении нового прецедента, определяется наиболее подходящая запись из хранилища с соответствующим решением, после чего данное решение проходит проверку. Если подобного прецедента не существует либо его решение неприемлемо, то система корреляции строит новое решение на основе старых. Полученное решение проверяется на корректность применения к прецеденту и в случае успеха вместе с ним добавляется в хранилище, иначе — ищется новое решение [3, 6, 7].

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

4.4. Метод на основе использования Байесовской сети. В основе данного метода лежит модель направленного ациклического графа. Суть метода заключается в расположении в вершинах сети корреляционных признаков, а связывающие их направленные дуги задают отношения условной независимости их значений [3, 4, 7, 11]. Обучение системы корреляции производится за счет последовательного вычисления значений условных вероятностей вершин, переменные которых неизвестны. Данная операция выполнима за счет подачи на

вход данных со значениями переменных корреляционных признаков. Метод подразумевает наличие обучающей выборки событий безопасности и является динамическим, самообучаемым и вероятностным. Путь прохождения результата вычисления на этапе работы Байесовской сети определяется только в случае сохранения множества корреляционных параметров в вершинах сети [7].

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

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

4.5. Метод на основе использования нейронной сети. Основой данного метода является математическая модель, состоящая из нейронов, имеющих собственное состояние, и линий связи (синапсов), определяющих влияние входных для него нейронов на данное состояние. Результатом выполнения работы каждого нейрона является аксон, значение которого может быть использовано в качестве входных нейронов для нейронов более высокого уровня [3, 4, 7, 10, 13]. В рамках корреляции событий безопасности в роли нейронов выступают корреляционные признаки. В любой схеме нейронной сети содержатся минимум 2 уровня — нулевой и единичный. На этапе обучения на входные нейроны поступают множества значений корреляционных признаков потока событий, при этом влияние входных нейронов на нейроны следующего уровня изначально задается случайно [7, 13]. По мере обучения системы значения влияний корректируются для соответствия заданному результату. Такой подход является обучением с учителем.

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

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

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

4.6. Комбинированные (гибридные) и другие методы корреляции. В реальных условиях применение только одного метода корреляции для анализа исследуемых инфраструктур недостаточно для получения точной оценки защищенности компьютерной сети и управления событиями и информацией безопасности [3, 7, 10]. Данное положение обусловлено рядом факторов, таких как: (1) вычислительная сложность метода; (2) функциональная нагрузка метода; (3) ресурсопотребление и др. Используя несколько методов на критичных этапах процесса корреляции, при пересечении множеств получаемых результатов возможно добиться более высокой точности оценки защищенности анализируемой инфраструктуры и определения текущей ситуации по компьютерной безопасности. Также возможен вариант последовательного применения разных методов корреляции на одном из этапов общего процесса корреляции. Данный факт связан с обработкой данных разных уровней, например, при анализе простых и более сложных событий безопасности.

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

5. Заключение. Данная статья завершает описание проведенного исследования по анализу методов корреляции событий безопасности в SIEM-системах. Были приведены основные научные работы по данному направлению, на основе которых была разработана классификация методов корреляции событий безопасности. В статье

подробно описывается каждый элемент классификации с примерами значений для конкретных методов. Был выделен и рассмотрен ряд методов, являющихся наиболее распространенными в существующих решениях. Данные методы были классифицированы, что позволило оценить их теоретическую эффективность на различных этапах процесса корреляции. Также рассмотрены методы обучения для выполнения оценки защищенности компьютерной сети с использованием разных методов корреляции. Целью дальнейших исследований является более глубокий анализ методов корреляции, разработка новых подходов к их реализации и экспериментальная оценка их эффективности.

Литература

1. *Kotenko I.V., Chechulin A.A.* A Cyber Attack Modeling and Impact Assessment Framework // Proceedings of 5th International Conference on Cyber Conflict 2013 (CyCon 2013). 2013. pp. 119–142.
2. *Kotenko I.V., Polubelova O.V., Saenko I.V.* The Ontological Approach for SIEM Data Repository Implementation // IEEE International Conference on Green Computing and Communications. IEEE Computer Society. 2012. pp. 761–766.
3. *Guerer D.W., Khan I., Ogler R., Keffer R.* An artificial intelligence approach to network fault management // SRI International. 1996. 10 p.
4. *Tiffany M.* A survey of event correlation techniques and related topics. URL: http://www.tiffman.com/netman/netman.html (дата обращения: 26.04.2016).
5. *Hasan M.* A conceptual framework for network management event correlation and filtering systems // Proceedings of the Sixth IFIP/IEEE International Symposium on Integrated Network Management. 1999. pp. 233–246.
6. *Hanemann A., Marcu P.* Algorithm Design and Application of Service-Oriented Event Correlation // Proceedings of Conference BDIM 2008, 3rd IEEE/IFIP International Workshop on Business-Driven IT Management. 2008. pp. 61–70.
7. *Muller A.* Event Correlation Engine. Master`s Thesis // Swiss Federal Institute of Technology Zurich. 2009. 165 p.
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. *Kruegel C., Valeur F., Vigna G.* Intrusion Detection and Correlation: Challenges and Solutions // University of California, Santa Barbara, USA: Springer. 2005. pp. 29-33.
10. *Ghorbani A.A., Lu W., Tavallaee M.* Network Intrusion Detection and Prevention // Springer. 2010. 224 p.
11. *Sadoddin R., Ghorbani A.* Alert Correlation Survey: Framework and Techniques // Proceedings of 2006 International Conference on Privacy, Security and Trust: Bridge the Gap Between PST Technologies and Business Services (PST`06). 2006. Article no. 37.
12. *Ning P., Xu D.* Correlation analysis of intrusion alerts // Intrusion Detection Systems: series Advances in Information Security. Springer. 2008. vol. 38. pp. 65–92.
13. *Elshoush H.T., Osman I.M.* Alert correlation in collaborative intelligent intrusion detection systems — A survey // Applied Soft Computing. 2011. pp. 4349–4365.
14. *Файзуллин Р.Р., Васильев В.И.* Метод оценки защищенности сети передачи данных в системе мониторинга и управления событиями информационной безопасности на основе нечеткой логики // Вестник УГАТУ. 2013. Вып. 17. № 2(55). С. 150–156.

15. *Zurutuza U., Uribeetxeberria R.* Intrusion Detection Alarm Correlation: A Survey // Proceedings of IADAT International Conference on Telecommunications and computer Networks. 2004. pp. 1–3.
16. *Jakobson G., Weissman M.D.* Alarm correlation // IEEE Network. 1993. vol. 7(6). pp. 52–59.

Федорченко Андрей Владимирович — младший научный сотрудник лаборатории проблем компьютерной безопасности, Федеральное государственное бюджетное учреждение науки Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов: безопасность компьютерных сетей, обнаружение вторжений, вредоносные программы. Число научных публикаций — 14. fedorchenko@comsec.spb.ru, http://comsec.spb.ru/ru/staff/fedorchenko; 14-я линия В.О., 39, ком. 205, Санкт-Петербург, 199178; р.т.: +7-812-328-71-81.

Левшун Дмитрий Сергеевич — программист лаборатории проблем компьютерной безопасности, Федеральное государственное бюджетное учреждение науки Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов: компьютерная безопасность, защита встроенных устройств, системы киберфизической безопасности, безопасность распределённых систем, корреляция событий безопасности. Число научных публикаций — 5. levshun@comsec.spb.ru, http://comsec.spb.ru/levshun; 14-я линия В.О., 39, ком. 205, Санкт-Петербург, 199178; р.т.: +7-(812)-328-71-81.

Чечулин Андрей Алексеевич — к-т техн. наук, старший научный сотрудник лаборатории проблем компьютерной безопасности, Федеральное государственное бюджетное учреждение науки Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов: безопасность компьютерных сетей, обнаружение вторжений, анализ сетевого трафика, анализ уязвимостей. Число научных публикаций — 150. andreych@bk.ru, http://comsec.spb.ru/ru/staff/chechulin; 14-я линия В.О., 39, ком. 205, Санкт-Петербург, 199178; р.т.: +78123287181.

Котенко Игорь Витальевич — д-р техн. наук, профессор, заведующий лабораторией проблем компьютерной безопасности, Федеральное государственное бюджетное учреждение науки Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов: безопасность компьютерных сетей, в том числе управление политиками безопасности, разграничение доступа, аутентификация, анализ защищенности, обнаружение компьютерных атак, межсетевые экраны, защита от вирусов и сетевых червей, анализ и верификация протоколов безопасности и систем защиты информации, защита программного обеспечения от взлома и управление цифровыми правами, технологии моделирования и визуализации для противодействия кибер–терроризму. Число научных публикаций — 450. ivkote@comsec.spb.ru, http://www.comsec.spb.ru; 14-я линия В.О., 39, Санкт-Петербург, 199178; р.т.: +7-(812)-328-71-81, Факс: +7(812)328–4450.

Поддержка исследований. Работа выполнена при финансовой поддержке РФФИ (проекты №14-07-00697, 14-07-00417, 15-07-07451, 16-37-00338, 16-29-09482 офи_м), при частичной поддержке бюджетных тем № 0073-2015-0004 и 0073-2015-0007, а также гранта РНФ 15-11-30029 в СПИИРАН.

DOI 10.15622/sp.49.11

A.V. Fedorchenko, D.S. Levshun, A.A. Chechulin, I.V. Kotenko

AN ANALYSIS OF SECURITY EVENT CORRELATION TECHNIQUES IN SIEM-SYSTEMS. PART 2

Fedorchenko A.V., Levshun DSL, Chechulin A.A., Kotenko I.V. **An Analysis of Security Event Correlation Techniques in SIEM-Systems. Part 2.**

Abstract. The paper proceeds research of the security event correlation methods in Security Information and Event Management (SIEM) systems. In this part we consider correlation methods of information security events that can be applied during separate correlation stages described in the previous paper. Classification of the considered correlation methods and analysis of their advantages and disadvantages are provided. The effectiveness of using these methods at different stages of the correlation process is evaluated.

Keywords: data correlation techniques; security event; security event analysis; computer network security evaluation systems; SIEM-systems.

Fedorchenko Andrey Vladimirovich — junior researcher of computer security problems laboratory, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Science (SPIIRAS). Research interests: computer network security, intrusion detection, malware. The number of publications — 14. fedorchenko@comsec.spb.ru, http://comsec.spb.ru/ru/staff/fedorchenko; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7-812-328-71-81.

Levshun Dmitry Sergeevich — software developer of computer security problems laboratory, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Science (SPIIRAS). Research interests: distributed system security, embedded devices, event correlation, cyber-physical security systems. The number of publications — 5. levshun@comsec.spb.ru, http://comsec.spb.ru/levshun; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7-(812)-328-71-81.

Chechulin Andrey Alexeevich — Ph.D., senior researcher of computer security problems laboratory, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Science (SPIIRAS). Research interests: computer network security, intrusion detection, analysis of the network traffic, vulnerability analysis. The number of publications — 150. andreych@bk.ru, http://comsec.spb.ru/ru/staff/chechulin; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +78123287181.

Kotenko Igor Vitalievich — Ph.D., Dr. Sci., professor, head of computer security problems Laboratory, St. Petersburg Institute for Informatics and Automation of Russian Academy of Sciences (SPIIRAS). Research interests: computer network security, including security policy management, access control, authentication, network security analysis, intrusion detection, firewalls, deception systems, malware protection, verification of security systems, digital right management, modeling, simulation and visualization technologies for counteraction to cyber terrorism. The number of publications — 450. ivkote@comsec.spb.ru, http://www.comsec.spb.ru; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7-(812)-328-71-81, Fax: +7(812)328–4450.

Acknowledgements. This research is supported by RFBR (projects No. 14-07-00697, 14-07-00417, 15-07-07451, 16-37-00338, 16-29-09482 офи_м), in part by the budget (projects No. 0073-2015-0004 and 0073-2015-0007) and by the grant of RSF 15-11-30029 in SPIIRAS.

References

1. Kotenko I.V., Chechulin A.A. A Cyber Attack Modeling and Impact Assessment Framework. Proceedings of 5th International Conference on Cyber Conflict 2013 (CyCon 2013). 2013. pp. 119–142.
2. Kotenko I.V., Polubelova O.V., Saenko I.V. The Ontological Approach for SIEM Data Repository Implementation. IEEE International Conference on Green Computing and Communications. IEEE Computer Society. 2012. pp. 761–766.
3. Guerer D.W., Khan I., Ogler R., Keffer R. An artificial intelligence approach to network fault management. SRI International. 1996. 10 p.
4. Tiffany M. A survey of event correlation techniques and related topics. Available at: http://www.tiffman.com/netman/netman.html (accessed: 26.04.2016).
5. Hasan M. A conceptual framework for network management event correlation and filtering systems. Proceedings of the Sixth IFIP/IEEE International Symposium on Integrated Network Management. 1999. pp. 233–246.
6. Hanemann A., Marcu P. Algorithm Design and Application of Service-Oriented Event Correlation. Proceedings of Conference BDIM 2008, 3rd IEEE/IFIP International Workshop on Business-Driven IT Management. 2008. pp. 61–70.
7. Muller A. Event Correlation Engine. Master`s Thesis. Swiss Federal Institute of Technology Zurich. 2009. 165 p.
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. Kruegel C., Valeur F., Vigna G. Intrusion Detection and Correlation: Challenges and Solutions. University of California. Santa Barbara. USA: Springer. 2005. pp. 29-33.
10. Ghorbani A.A., Lu W., Tavallaee M. Network Intrusion Detection and Prevention. Springer. 2010. 224 p.
11. Sadoddin R., Ghorbani A. Alert Correlation Survey: Framework and Techniques. Proceedings of 2006 International Conference on Privacy, Security and Trust: Bridge the Gap Between PST Technologies and Business Services (PST`06). 2006. Article no. 37.
12. Ning P., Xu D. Correlation analysis of intrusion alerts. *Intrusion Detection Systems: series Advances in Information Security.* Springer, 2008. vol. 38. pp. 65–92.
13. Elshoush H.T., Osman I.M. Alert correlation in collaborative intelligent intrusion detection systems — A survey. *Applied Soft Computing.* 2011. pp. 4349–4365.
14. Fajzullin R.R., Vasil'ev V.I. [Protectability assessment method of a data-transmission network in security information and event management system on a basis of fuzzy logic]. *Vestnik UGATU — Proceedings USATU*. 2013. vol. 17. no. 2(55). pp. 150–156. (In Russ.).
15. Zurutuza U., Uribeetxeberria R. Intrusion Detection Alarm Correlation: A Survey. Proceedings of IADAT International Conference on Telecommunications and computer Networks. 2004. pp. 1–3.
16. Jakobson G., Weissman M.D. Alarm correlation. *IEEE Network*. 1993. vol. 7(6). pp. 52-59.