УДК 004.056
doi:10.31799/1684-8853-2019-2-57-67

Научные статьи
Articles

Открытые задачи визуального анализа в системах управления информационной безопасностью

Е. С. Новикова[а, б], *канд. техн. наук, доцент, orcid.org/0000-0003-2923-4954, novikova@comsec.spb.ru*
И. В. Котенко[а], *доктор техн. наук, профессор, orcid.org/0000-0001-6859-7120*
[а]*Санкт-Петербургский институт информатики и автоматизации РАН, 14-я линия В. О., 39, Санкт-Петербург, 199178, РФ*
[б]*Санкт-Петербургский государственный электротехнический университет «ЛЭТИ», Профессора Попова ул., 5, Санкт-Петербург, 197376, РФ*

Введение: *методики визуального анализа позволяют эффективно исследовать постоянно увеличивающиеся объемы данных, генерируемые сенсорами безопасности, и способствуют своевременному и обоснованному реагированию на угрозы. В современных системах управления информационной безопасностью реализованы различные решения по обработке больших объемов данных и интеграции различных источников, которые могут быть использованы для построения полноценной системы визуального анализа инцидентов безопасности.* ***Цель:*** *исследование методик визуального анализа, реализованных в системах управления информационной безопасностью и предназначенных для исследования событий безопасности в контексте основных задач визуального анализа, включая верификацию корректности работы моделей автоматического анализа данных.* ***Результаты:*** *выявлено существующее противоречие между возможностями систем управления информационной безопасностью по анализу данных и реализацией этих возможностей. Практически отсутствуют методики визуального анализа данных, поддерживающие визуальную корреляцию данных от сенсоров безопасности и визуальную валидацию автоматических моделей анализа, позволяющую оценить эффективность и корректность их функционирования. Возможным решением этого противоречия является применение методик, реализующих гибкий механизм настройки анализируемых атрибутов событий сетевых устройств и сенсоров. В работе рассмотрены основные подходы к построению таких методик, обсуждаются их достоинства и недостатки. Предложена панель управления, предназначенная для мониторинга поведения автоматической модели анализа сетевого трафика в системе облачных вычислений, которая позволяет контролировать функционирование модели анализа, выполнять визуальную корреляцию исследуемых параметров и отслеживать характер изменений в сетевых потоках.* ***Практическая значимость:*** *результаты работы могут быть использованы при проектировании инструментов визуального анализа для исследования событий безопасности, мониторинга потоков данных и поведения автоматических моделей анализа.*

Ключевые слова *— визуальная аналитика, управление информационной безопасностью, визуальная корреляция данных, визуальная валидация моделей анализа, матричное представление данных.*

Для цитирования: Новикова Е. С., Котенко И. В. Открытые задачи визуального анализа в системах управления информационной безопасностью. *Информационно-управляющие системы*, 2019, № 2, с. 57–67. doi:10.31799/1684-8853-2019-2-57-67
For citation: Novikova E. S., Kotenko I. V. Open challenges in visual analytics for security information and event management. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2019, no. 2, pp. 57–67 (In Russian). doi:10.31799/1684-8853-2019-2-57-67

Введение

В настоящее время методики *визуальной аналитики* (visual analytics), в которых сочетаются достоинства графического представления многомерных данных и их автоматизированного анализа, считаются едва ли не единственным способом анализа и исследования данных большого объема [1]. Визуальный анализ данных, описывающих безопасность информационной системы, является относительно новым, но активно развивающимся и востребованным направлением визуальной аналитики [2]. Методы визуального анализа помогают эффективно исследовать постоянно увеличивающиеся объемы данных, генерируемые сенсорами безопасности и, как следствие, способствуют своевременному и обоснованному реагированию на угрозы.

Ниже приведены задачи информационной безопасности, для решения которых могут быть использованы методики визуальной аналитики: оперативный контроль периметра компьютерной сети; оценка уровня защищенности компьютерной сети; обнаружение внутренних нарушителей; исследование инцидентов безопасности и формирование сценариев атаки; верификация политик безопасности сетевых устройств; исследование вредоносного кода; верификация автоматических моделей анализа, используемых для управления информационной безопасностью.

Выделим шесть основных проблем проектирования методик визуального анализа: большой объем данных; многообразие источников данных; отсутствие связи (синхронизации) между источниками данных; низкое качество данных; необходимость формирования паттернов нор-

мального поведения информационной системы; сложность отслеживания развития информационных угроз и реагирования [3, 4].

Возможные решения по обработке больших объемов данных и интеграции различных источников для построения компонента визуального анализа систем управления информационной безопасностью могут быть найдены в технологии управления информацией и событиями безопасности (Security Information and Event Management — SIEM), которая дает возможность собирать и обрабатывать данные от различных сенсоров безопасности, например, систем обнаружения/предотвращения вторжений, сканеров уязвимости, межсетевых экранов, роутеров и т. д., решая указанные проблемы.

В настоящей работе рассмотрены основные методики визуального анализа, применяемые в системах управления информацией и событиями безопасности, выявлены основные открытые задачи и определены возможные пути их решения. Отличительной особенностью данной работы от других исследований [2, 5, 6] является оценка методик визуализации существующих SIEM-систем в контексте основных задач визуальной аналитики: мониторинг текущей ситуации (оперативный контроль); исследование данных (исторический анализ данных); формирование отчетов; верификация корректности работы моделей автоматического анализа [7]. В работе также представлен макет панели управления, предназначенной для валидации модели анализа, используемой для обнаружения атак в системе облачных вычислений OpenStack, который построен с помощью выявленных методик визуального анализа многомерных данных.

Обзор методик визуального анализа данных в SIEM-системах

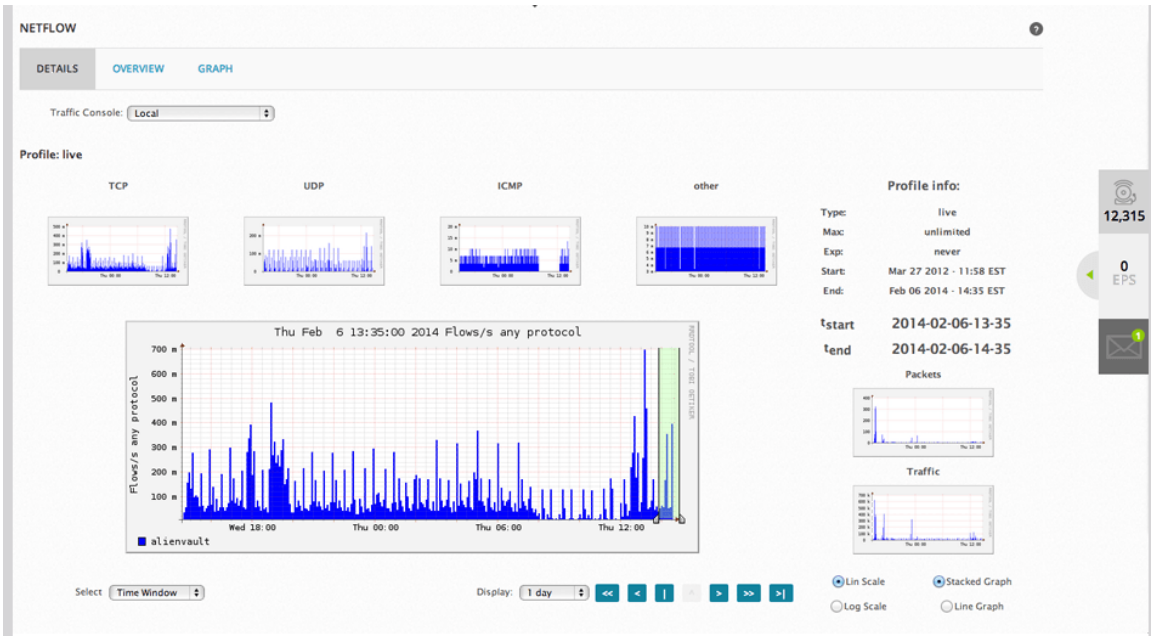
Помимо широких возможностей по обработке больших потоков разнородных данных, SIEM-системы позволяют выполнять исторический анализ событий безопасности и их последствий, представлять трассы атак [8], моделировать и формировать новые правила обнаружения информационных угроз [9] и оценки уровня их критичности [10]. Они предоставляют механизмы по управлению вычислительными ресурсами информационной системы, выполняют ее проверку на соответствие нормативным документам в области безопасности и создают отчетную документацию [10].

Для решения этих задач в SIEM-системах используются так называемые *панели управления* (dashboards), которые задают способ организации информации, необходимой для решения одной или нескольких задач в графическом виде, обеспечивающем ее понимание с первого взгляда [11]. В зависимости от характера выполняемой задачи можно выделить *аналитические* и *оперативные* панели управления. Аналитические панели, как правило, содержат обобщающие результаты обработки и анализа данных о событиях, в то время как оперативные панели предоставляют низкоуровневые, часто необработанные данные, например записи межсетевого экрана.

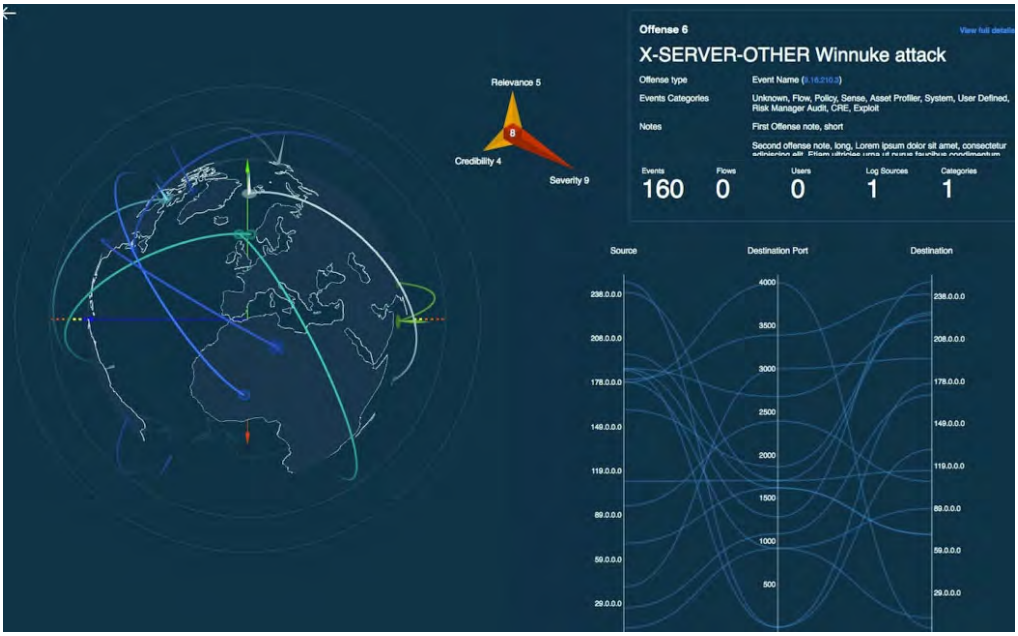
Оперативные панели управления, используемые для мониторинга событий безопасности или просмотра данных от сенсоров безопасности в режиме реального времени, чаще всего включают в себя табличное представление данных, в некоторых случаях дополненное моделями визуализации — столбчатыми диаграммами и линейными графиками. Например, на рис. 1 [12] представлена информация от сенсора учета сетевого трафика NetFlow. Очевидно, что наиболее интересными для аналитика характеристиками сетевого трафика являются входящий/исходящий объем сетевого трафика, его распределение по используемым протоколам, поэтому ключевым элементом данной панели является множество линейных графиков, показывающих изменение сетевого трафика во времени с учетом сетевого протокола.

Можно выделить два типа аналитических панелей управления: 1) панели, представляющие статистические данные, характеризующие функционирование SIEM-системы и ее компонентов на достаточно высоком уровне, и 2) панели управления, предназначенные для исследования инцидентов безопасности. Показательным примером аналитических панелей управления первого типа являются панели управления, отображающие статистику по сенсорам безопасности. Они обычно характеризуют количественную составляющую их функционирования, т. е. общее число срабатываний правил безопасности (сигналов тревоги), распределение срабатываний по типам (правилам), выявление объектов сети, наиболее подверженных атакам, основных источников атак и т. д. Для представления этих данных используются двухмерные стандартные модели визуализации (столбчатые или линейные диаграммы, круговые диаграммы, линейные графики с временной шкалой) и текстовые представления (таблицы).

Исследование инцидентов безопасности обычно выполняется с помощью интерактивных таблиц, элементы которых часто реализуются в виде гиперссылок, по которым осуществляется переход к другим данным, благодаря чему значительно упрощается поиск требуемой информации. Некоторые SIEM-системы используют более сложные модели визуализации для анализа событий безопасности и определения возможного



■ ***Рис. 1.*** Панель просмотра данных от NetFlow в SIEM-системе OSSIM
■ ***Fig. 1.*** The dashboard for monitoring Netflow data in the OSSIM SIEM



■ ***Рис. 2.*** Панель исследования событий безопасности в подключаемом компоненте QRadar Pulse SIEM-системы IBM QRadar
■ ***Fig. 2.*** The dashboard for security incident investigation in the QRadar Pulse, offense visualization application for the QRadar SIEM

сценария атак. Так, например, в SIEM-системе IBM QRadar для исследования распределения атак различного типа в пространстве предложена трехмерная модель земного шара, на который наносятся события разного типа (рис. 2) [13].

Для представления событий используется сложный глиф, характеризующий следующие аспекты: уровень критичности событий (высота графического элемента на глобусе); частоту событий (размер круга-основания графического

элемента на круге); тенденцию частоты новых событий (концентрические круги вокруг графического элемента). Связь между событиями показана дугами, соединяющими исходный IP-адрес и IP-адрес назначения. Детали событий безопасности отображаются с помощью графа на параллельных координатах, оси которого достаточно традиционны для исследования сетевых потоков: IP-адрес источника, номер порта получателя и IP-адрес получателя.

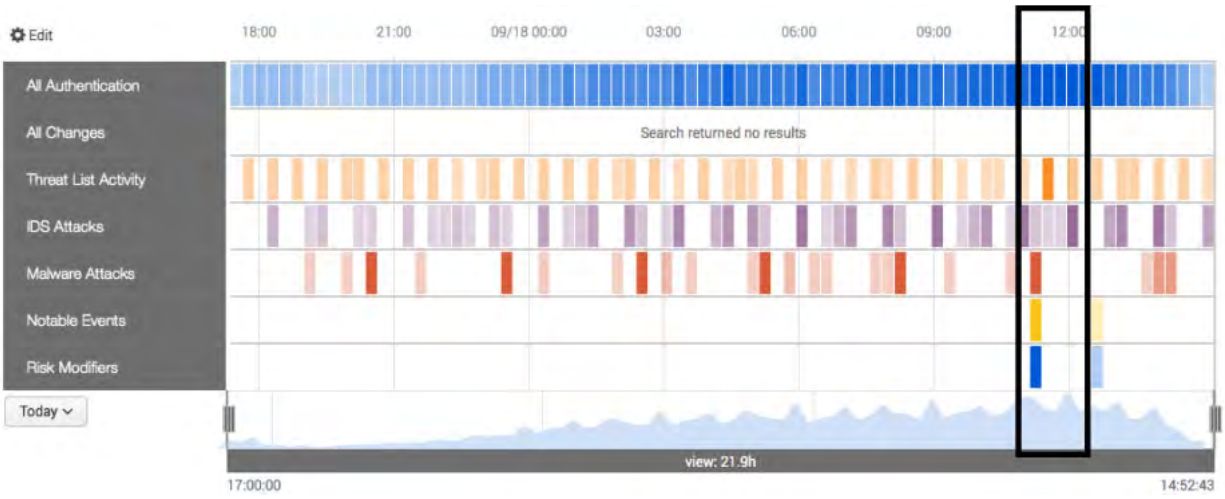
В ArcSight ESM имеется возможность представить последовательность событий в виде связного графа, вершинами которого являются внешние и внутренние хосты компьютерной сети и сетевые порты, ребра графа связывают атакующий и атакуемый хосты. Цветом указывается статус хоста — источник атаки (красный цвет), атакующий хост, захваченный в процессе атаки (голубой цвет), и атакуемый на текущий момент хост (белый цвет).

Интересная модель визуализации для анализа событий безопасности одного хоста на основе матричного представления предложена в системе Splunk (рис. 3) [14].

События безопасности сгруппированы по сенсорам безопасности, которые их генерируют: события системы аутентификации пользователей, события системы обнаружения вторжений, значимые события, определяемые по правилам пользователей, события антивирусных решений и т. д. События отображаются по оси *Y*. Тип события на графике кодируется цветом, причем интенсивность цвета указывает число событий в заданный период времени. По оси *X* откладывается время. Внизу матричного представления расположен обычный линейный график, который отображает общее количество событий во времени. Отличительными характеристиками данной модели визуализации являются возможность установить временные связи между событиями разного типа и ее относительная простота. Например, на рис. 3 выделен интервал времени, которому соответствует максимальное число событий всех типов. При этом легко установить, что сначала от системы обнаружения вторжений было зафиксировано большое число событий, за которыми последовали события, классифицированные как проявление вредоносной активности. Они обозначены как значимые события, связанные с изменением уровня риска хоста. Затем были выявлены события, входящие в список информационных угроз.

Исследование SIEM-систем, являющихся лидерами рынка систем управления информационной безопасностью [15], показало, что основными методиками визуализации в SIEM-системах являются: интерактивные таблицы; стандартные двухмерные и трехмерные методики визуализации, представленные столбчатыми и линейчатыми диаграммами, линейными графиками с временной шкалой и круговыми диаграммами; географические карты, в том числе трехмерные, используемые в сочетании с пиктограммами для отображения обобщенных показателей безопасности; связные графы, также часто используемые наряду с пиктограммами для описания статуса объекта, обозначенного вершиной графа.

Стандартные двухмерные модели визуализации в основном отображают различную статистическую информацию. Связные графы применяются для отображения топологии компьютерных сетей, информационных ресурсов сети, информационных потоков между узлами [12–14, 16]. Они также используются для графического



■ ***Рис. 3.*** Визуальный анализ событий безопасности одного хоста
■ ***Fig. 3.*** Visualization of the security events per computer network asset

представления связей между событиями безопасности, действиями пользователей в системе. Практически во всех SIEM-системах реализована поддержка географических карт для отображения географического расположения хостов контролируемой сети и событий безопасности. С их помощью пользователь имеет возможность отслеживать развитие атаки в пространстве. Кроме того, они помогают системному администратору определять границы ответственности подразделений организации.

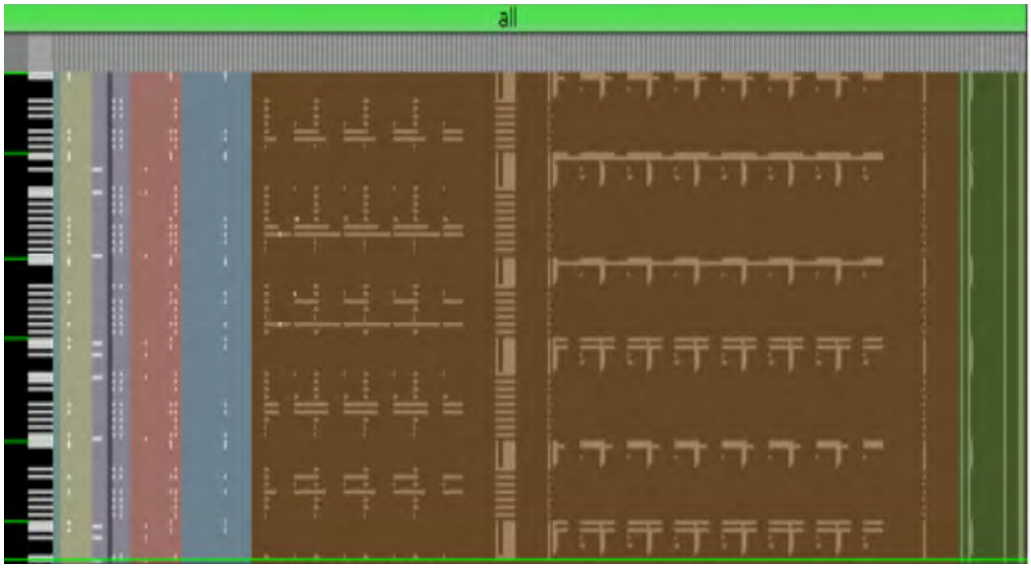
Перечисленные выше методики визуализации предназначены для представления одномерных и двухмерных данных, легко воспринимаются человеком, что делает их полезными для выполнения задач оперативного контроля и представления результатов исследования. Исследование событий безопасности в большинстве случаев заключается в изучении данных, представленных в разных таблицах, лишь в некоторых случаях имеется возможность визуально проанализировать события безопасности, построив граф событий. Вместе с тем именно SIEM-системы дают возможность построить полноценную систему визуального анализа состояния информационной безопасности благодаря уже реализованным механизмам сбора и подготовки данных от разнородных источников данных. Таким образом, очевидна необходимость использования в SIEM-системах моделей графического представления многомерных данных, которые позволяют выявлять скрытые зависимости как между атрибутами объектов, так и между данными от разных источников, что в конечном итоге позволит обнаруживать новые сценарии атак.

Определенные попытки для устранения этой проблемы предприняты ведущими поставщиками SIEM-систем, которые выпустили подключаемые модули, реализующие механизмы визуального анализа для исследования событий безопасности [13–15]. Однако предложенные механизмы визуального анализа опираются на определенные атрибуты сетевого трафика или событий безопасности, которые показали свою состоятельность при обнаружении традиционных сетевых атак и, соответственно, широко используются в системах обнаружения вторжений (СОВ). В связи с этим необходимость в использовании предложенных методик визуального анализа резко снижается, поскольку они, по сути, дублируют результат работы системы обнаружения вторжений. Одним из исключений является SIEM-система Splunk, в которой именно пользователь определяет исходные данные для анализа, задает модель их графического представления и методики взаимодействия с нею. Однако подобная свобода действий предполагает наличие у пользователей опыта по проектированию моделей визуального анализа, панелей управления, а также навыков по разработке программ. Очевидно, что подобными навыками обладают далеко не все специалисты по информационной безопасности.

Вызовы и возможные решения

Анализ существующих методик обозначил противоречия между возможностями SIEM-систем по анализу данных и реализацией этих возможностей. С одной стороны, SIEM-системы являются удобной платформой для построения системы визуального анализа благодаря масштабируемым и эластичным технологиям сбора, предварительной обработке данных от разнородных источников, а особенности архитектурного решения SIEM-систем позволяют использовать подключаемые компоненты, реализующие методики визуального анализа. С другой стороны, представленные методики визуального анализа в основном предназначены для выполнения задач оперативного контроля и представления результатов. Практически отсутствуют методики визуального анализа, поддерживающие исследование событий безопасности, визуальную корреляцию данных, визуальную валидацию автоматических моделей анализа, позволяющую оценить эффективность и корректность их функционирования. Возможным решением этой проблемы является применение методик, осуществляющих анализ журналов сетевых устройств и сенсоров безопасности независимо от их природы и без привязки к значениям определенных атрибутов [4, 17–21]. Это позволяет обнаруживать качественно новые атаки и взаимосвязи между событиями. Данные методики можно условно разделить на две группы: 1) построение универсального графического анализатора журнала сетевых сенсоров безопасности и сетевых устройств, который выполняет анализ логов по некоторому множеству правил-фильтров, заданному пользователем [4, 19]; 2) выполнение визуальной корреляции атрибутов логов, генерируемых несколькими сенсорами безопасности, причем атрибуты определяются пользователем, исходя из содержимого логов устройства [18, 20].

Примером первого подхода является методика, представленная в работе [21]. Сырые сетевые пакеты, поступающие от различных сенсоров сетевого трафика и информационной безопасности, преобразуются в сообщения, имеющие иерархическую структуру и заданные с помощью XML-подобного языка Packet Detail Markup Language (PDML), который описывает содержимое пакета, учитывая различные атрибуты сетевых протоколов. Пакеты отображаются вдоль оси Y, а атрибуты сетевых па-



■ ***Рис. 4.*** Пиксель-ориентированное представление логов сетевых устройств
■ ***Fig. 4.*** Pixel-oriented visualization of the security devices' logs

кетов — вдоль оси *X*. Таким образом, каждое сообщение представляет собой последовательность пикселей. Яркость (*i*, *j*) пикселя обозначает частоту атрибута *j* в сообщении *i*. На рис. 4 представлена предложенная авторами модель визуализации логов сетевых устройств [21].

Важным моментом в использовании данного подхода является корректный выбор цветовой схемы кодирования значения параметра, отображаемого с помощью пикселя. Она должна позволять акцентировать внимание аналитика на аномалии в данных, например, на редко встречаемых и часто встречаемых элементах, при этом следует помнить, что редко встречаемые значения не всегда являются признаком аномальной активности. Следует также учитывать, что из-за мелких размеров пиксели разного оттенка, но одинаковой степени насыщенности, плохо различаются человеком [22], и в результате для анализа создаваемого графического представления аналитику приходится прилагать значительные усилия. Другим очевидным недостатком этого подхода является возможность анализировать данные только от одного источника, что не позволяет выявлять связи между показаниями разных источников.

Вторая группа инструментов дает возможность визуально оценивать совокупность логов от разных устройств, а анализ выполняется по множеству атрибутов записей журналов, выбираемых самим пользователем. Для их представления используются достаточно простые модели визуализации — линейные графики, графики рассеивания, столбиковые диаграммы. В работе [18] «привязка» событий от разных источников данных осуществляется путем использования линейных графиков с общей временной шкалой. В [20] формируются временные срезы, в которых состояние информационной системы неизменно относительно параметров, заданных пользователями (число событий безопасности, число получателей/отправителей сетевых потоков, число запущенных программ и т. д.). Для каждого такого среза строится множество моделей визуализации, отражающих основные характеристики потоков данных за данный период времени. Таким образом, несмотря на то, что явной привязки событий к шкале времени нет, имеется возможность оценить общий характер потоков и увидеть определенные связи между их атрибутами. Использование временных срезов как средства агрегации характеристик потоковых данных может быть рассмотрено как механизм снижения когнитивной нагрузки на аналитика при анализе высокоскоростных потоков данных большого объема: графическое представление должно быть достаточно предсказуемым и постоянным, не зависящим от скорости, объема и разнообразия исходных данных. В качестве возможных недостатков предложенного подхода можно выделить использование достаточно сложных моделей визуализации (плоскостных карт деревьев, связных графов), которые могут отличаться от одного временного среза к другому, что требует их тщательного изучения со стороны аналитика. Кроме того, модели визуализации отображают данные, представленные только в текущем срезе потока, что не позволяет оценить изменения параметров потока данных в контексте состояния всей информационной системы.

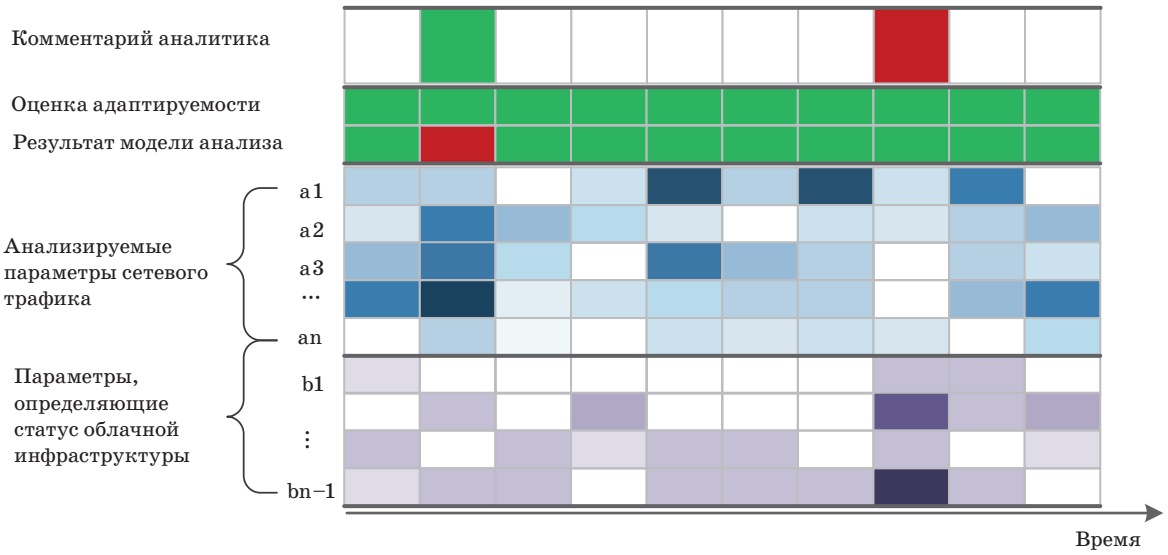
Панель управления автоматической моделью анализа сетевого трафика в инфраструктуре облачных вычислений

Задача визуальной валидации автоматических моделей анализа сетевого трафика и выявления аномалий связана с предоставлением механизмов мониторинга безопасности, используемых для выявления различных информационных угроз. Они должны позволять пользователю отслеживать результаты работы моделей анализа, оценивать уровень ее адаптируемости к изменениям в потоках данных и возможности переобучения и, в случае необходимости, отмечать ситуации ложных срабатываний, необходимые для последующего анализа и настройки или доработки.

Исходя из этих требований была разработана панель управления, предназначенная для мониторинга функционирования модели анализа сетевого трафика, применяемой для обнаружения атак в инфраструктуре облачных вычислений [23]. В ее основе лежат модели классификации. Механизм переобучения моделей, адаптирующий систему защиты к изменениям структуры и объема трафика в системе облачных вычислений, основан на оценке параметра уверенности (confidence) обученной модели в предсказании типа трафика. Авторы предлагают исследовать результат ее работы с учетом параметров сетевого трафика, на основе которых принимается решение о его легитимности, внутренней оценки адаптируемости модели анализа к изменениям в сетевом трафике, а также параметров, описывающих общий статус облачной инфраструктуры. Первые две группы параметров фиксированы, а последняя задается пользователем системы обнаружения атак в системе облачных вычислений и предназначена для выявления косвенных признаков наличия необнаруженной вредоносной активности в системе.

Ключевым элементом разработанной панели управления является матричное представление данных (рис. 5). Выбор данной модели обусловлен тем фактом, что она позволяет выполнять визуальную корреляцию данных, а также выявлять определенные паттерны в них, например, изменения в сетевом трафике, периодически повторяющиеся во времени. Контролируемые параметры откладываются по оси *Y*, а время — по оси *X*. Во время функционирования модели каждые 5 секунд формируются векторы, характеризующие сетевой трафик. Однако анализировать такой объем информации затруднительно даже визуально, поэтому было решено отображать интервалы времени, в течение которых в данных не происходит значительных изменений, и модель анализа классифицирует их одинаково. Если характер данных меняется сильно, то формируется новый срез данных.

Для определения значимости изменений в данных и необходимости формировать новый срез данных используются принципы кластеризации на основе оценки плотности распределения объектов в пространстве. Анализируемый моделью вектор данных рассматривается как точка в многомерном пространстве, при появлении новой точки вычисляются расстояния до множества точек, уже входящих в текущий срез данных, и если максимальное расстояние не превышает некоторого порогового значения ε, задаваемого пользователем, то точка включается



■ ***Рис. 5***. Представление данных для валидации модели анализа трафика
■ ***Fig. 5***. Data visualization for traffic analysis model validation

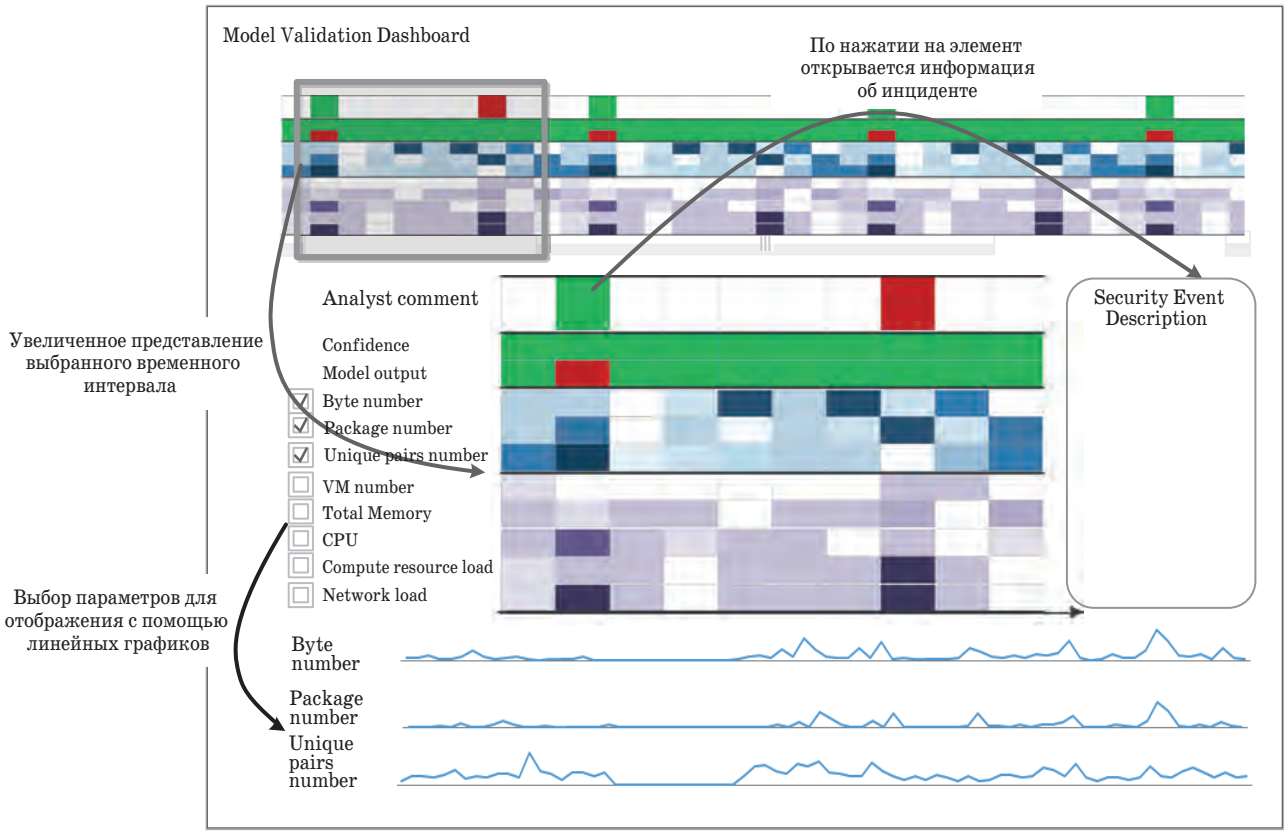
в текущий срез данных, в противном случае формируется новый срез. В качестве метрики расстояния применяется евклидова метрика. Новый срез данных формируется также в том случае, если меняется результат анализа сетевых потоков.

Откладываемые интервалы времени по шкале X могут иметь разную длительность, несмотря на то, что визуально они одинаковы. Элементом матрицы является среднее значение параметра, вычисленное для множества значений заданного среза данных. Полученное значение кодируется насыщенностью цвета. Чем насыщенней цвет элемента, тем выше значение параметра. Значение насыщенности цвета формируется путем нормализации значений в пределах интервала [0, 1]. Максимальное и минимальное значения параметра определяются или за весь исследуемый период времени (при использовании модели визуализации для исторического анализа данных), или за последний заданный интервал времени, например неделю, при мониторинге данных в режиме реального времени. Оттенок цвета используется для обозначения параметров, принадлежащих к одной группе.

Таким образом, пользователь, визуально сопоставляя значения параметров, генерируемых различными источниками, имеет возможность выявить скрытые взаимосвязи между ними, построить графические паттерны функционирования облачных сервисов. Аналитик может также обнаружить тенденции в изменении сетевого трафика и выявить подозрительные ситуации, которые не были распознаны как подозрительные.

Отдельная цветовая схема применяется для представления результатов модели анализа и результатов исследования инцидентов, выполненных аналитиком вручную. Результат модели анализа кодируется следующим образом. Если трафик классифицирован как легитимный, то соответствующий элемент матрицы окрашен зеленым цветом. В противном случае элемент матрицы имеет красный цвет. Для обозначения уровня уверенности модели также используется двухцветная палитра: красным цветом кодируются значения ниже порогового значения, поскольку в этом случае инициируется переобучение модели, зеленым цветом — значения выше порогового. Результаты ручного анализа кодируются тоже зеленым и красным цветами: зеленым, если результат анализа аналитика совпадает с решением модели анализа, красным, если не совпадает.

Построенная модель визуализации позволяет анализировать характер изменений в сетевом трафике, выявлять периодические паттерны функ-



■ ***Рис. 6***. Макет панели визуальной валидации модели анализа сетевого трафика
■ ***Fig. 6***. Dashboard prototype for visual validation of the network analysis model

ционирования инфраструктуры облачных вычислений, выполнять визуальную корреляцию данных, отслеживать эффективность работы модели анализа, отображать результаты исследования инцидентов безопасности и сравнивать их с решениями, полученными на базе модели анализа.

Недостатком предложенной модели визуализации является нелинейность шкалы времени из-за особенности формирования срезов данных. Для устранения этого недостатка было предложено использовать дополнительные модели визуализации. При их выборе авторы руководствовались тем, что специалисты по информационной безопасности предпочитают работать со стандартными двухмерными графическими представлениями (линейными графиками, гистограммами) [24], поэтому детальное представление данных осуществляется с помощью системы линейных графиков с общей шкалой времени. Они отображают изменения выбранных параметров во времени.

Матричное представление и система линейных графиков связаны между собой набором механизмов взаимодействия. На рис. 6 показан макет панели управления, поддерживающей процесс визуальной валидации модели анализа.

В верхней части панели управления расположено матричное представление параметров, оно формирует общую картину того, каким образом функционирует автоматическая модель анализа сетевых потоков. Кроме того, оно может быть использовано для мониторинга в режиме реального времени. В этом случае новые значения параметров добавляются справа. Поскольку данная модель является достаточно мелкой, для работы с ней предложен механизм увеличения, который позволяет выбрать часть матрицы и увеличить ее. Увеличенная часть матрицы находится в центре панели управления, и именно с ней в основном взаимодействует аналитик. Первая строчка матрицы является редактируемой, аналитик может вносить свои результаты исследования инцидентов, отмечая достоверность результатов автоматической модели анализа. Кроме того, он может записывать свои комментарии в текстовую форму «*Security Event Description*». Параметры, отображаемые с помощью линейных графиков, также задаются пользователем, они выбираются из множества параметров, для которых строится матричное представление.

Описанная панель управления реализована в системе управления облачной инфраструктуры OpenStack для мониторинга модели обнаружения атак в виде подключаемого компонента интерфейса пользователя Horizon.

Заключение

В статье исследованы известные методики визуального анализа, представленные в системах управления безопасностью, в частности в SIEM-системах, для исследования событий безопасности. Выделены противоречия между возможностями систем, связанных с механизмами подготовки исходных для анализа данных, и практической реализацией существующих решений.

Рассмотрены возможные пути решения представленных проблем, причем для каждой группы решений показаны потенциальные достоинства и недостатки, что позволит разработчикам в сфере информационной безопасности выбрать методику визуального анализа данных, подходящую для решения конкретной задачи. В работе также представлена панель управления, предназначенная для валидации автоматической модели выявления атак в инфраструктуре облачных вычислений, которая была разработана с учетом рассмотренных методик построения систем визуального анализа.

Дальнейшее направление исследований связано с выполнением оценки эффективности предложенной модели визуализации, ее адаптацией для контроля других моделей анализа, применяемых для решения различных задач информационной безопасности, а также с уточнением механизмов формирования срезов данных.

Работа выполнена при частичной финансовой поддержке РФФИ (проекты 16-29-09482, 18-37-20047, 18-07-01488 и 18-29-22034) и бюджетной темы 0073-2019-0002.

Литература

1. **Marty R.** *Visual Analytics — Delivering Actionable Security Intelligence*. https://www.blackhat.com/us-16/training/visual-analytics-delivering-actionable-security-intelligence.html (дата обращения: 16.01.2019).
2. **Guimarães V. T., Freitas C. M. D. S., Sadre R., Tarouco L. M. R., Granville L. Z.** A survey on information visualization for network and service management. *IEEE Communications Surveys & Tutorials*, 2016, vol. 18, no. 1, pp. 285–323. doi:10.1109/COMST.2015.2450538
3. **Best D. M., Endert A., Kidwell D.** 7 key challenges for visualization in cyber network defense. *Proc. of the Eleventh Workshop on Visualization for Cyber Security (VizSec 2014)*, ACM, 2014, pp. 33–40. doi:10.1145/2671491.2671497
4. **Stange J.-E., Dörk M., Landstorfer J., and Wettach R.** Visual filter: graphical exploration of network security log files. *Proc. of the Eleventh Workshop on Visualization for Cyber Security (VizSec 2014)*, ACM, 2014, pp. 41–48. doi:10.1145/2671491.2671503

5. **Котенко И. В., Новикова Е. С.** Визуальный анализ защищенности компьютерных сетей. *Информационно-управляющие системы*, 2013, № 3, с. 56–61.
6. **Zhang T., Wang X., Li Z.,** et al. A survey of network anomaly visualization. *Sci. China Inf. Sci.*, 2017, vol. 60, pp. 121101. doi:10.1007/s11432-016-0428-2
7. **Krause J., Perer A., Bertini E.** Using visual analytics to interpret predictive machine learning models. *Proc. of the 2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016)*, New York, NY, USA. 2016. https://arxiv.org/abs/1606.05685 (дата обращения 16.01.2019).
8. **Котенко Д. И., Котенко И. В., Саенко И. Б.** Методы и средства моделирования атак в больших компьютерных сетях: состояние проблемы. *Тр. СПИИРАН*, 2012, № 3 (22), с. 5–30.
9. **Котенко И. В., Степашкин М. В.** Анализ защищенности компьютерных сетей на основе моделирования действий злоумышленников и построения графа атак. *Тр. Института системного анализа Российской академии наук*, 2007, т. 31, с. 126–207.
10. **Котенко И. В., Саенко И. Б.** Архитектура системы интеллектуальных сервисов защиты информации в критически важных инфраструктурах. *Тр. СПИИРАН*, 2013, № 1 (24), с. 21–40.
11. **Few S.** *Information dashboard design the effective visual communication of data*. N. Y., O'Reilly, 2006. 224 p.
12. *SIEM-система OSSIM*. https://www.alienvault.com/products/ossim (дата обращения: 16.01.2019).
13. *Компонент визуальной аналитики Qradar Pulse SIEM-системы IBM Radar*. https://www.ibm.com/support/knowledgecenter/en/SS42VS_7.2.8/com.ibm.apps.doc/c_Qapps_Pulse_intro.html (дата обращения: 16.01.2019).
14. *SIEM-система Splunk Enterprise*. https://www.splunk.com/ru_ru/products/ splunk-enterprise.html (дата обращения: 16.01.2019).
15. *Gartner Magic Quadrant for Security Information and Event Management*. https://www.gartner.com/doc/3894573 (дата обращения: 16.01.2019).
16. *Компонент визуальной аналитики Interactive Discovery SIEM-системы HP ArcSight*. https://community.softwaregrp.com/t5/ArcSight-Interactive-Discovery/tkb-p/arcsight-interactive-discovery (дата обращения: 16.01.2019).
17. **Walton S., Maguire E., and Chen M.** Multiple queries with conditional attributes (QCATs) for anomaly detection and visualization. *Proc. of the Eleventh Workshop on Visualization for Cyber Security (VizSec 2014)*, ACM, 2014. doi:10.1145/2671491.2671502
18. **Humphries C., Prigent N., Bidan C., and Majorczyk F.** CORGI: combination, organization and reconstruction through graphical interactions. *Proc. of the Eleventh Workshop on Visualization for Cyber Security (VizSec 2014)*, ACM, 2014, pp. 57–64. doi:10.1145/2671491.2671494
19. **Landstorfer J., Herrmann I., Stange J., Dork M., and Wettach R.** Weaving a carpet from log entries: A network security visualization built with cocreation. *Proc. of the Visual Analytics Science and Technology (VAST)*, Paris, 2014, pp. 73–82. doi:10.1109/VAST.2014.7042483
20. **Fischer F., Keim D. A.** NstreamAware: real-time visual analytics for data streams to enhance situational awareness. *Proc. of the Eleventh Workshop on Visualization for Cyber Security (VizSec 2014)*, ACM, 2014, pp. 65–72. doi:10.1145/2671491.2671495
21. **Cappers B. C. M., van Wijk J.** SNAPS: Semantic network traffic analysis through projection and selection. *Proc. of the IEEE Symposium on Visualization for Cyber Security (VizSec 2015)*, 2015, pp. 1–8. doi:10.1109/ VIZSEC.2015. 7312768
22. **Stone M.** In color perception, size matters. *IEEE Computer Graphics and Applications*, 2012, vol. 32(2), pp. 8–13. doi:10.1109/MCG.2012.37
23. **Borisenko K., Smirnov A., Novikova E., and Shorov A.** DDoS attacks detection in cloud computing using data mining techniques. *Proc. of the Industrial Conf. on Data Mining (ICDM'2016)*, LNAI, vol. 9728, pp. 1–15. doi:10.1007/978-3-319-41561-1_15
24. **McKenna S., Staheli D., Fulcher C., Meyer M.** BubbleNet: A cyber security dashboard for visualizing patterns. *Computer Graphics Forum*, 2016, vol. 35, pp. 281–290. doi:10.1111/cgf.12904

UDC 004.056
doi:10.31799/1684-8853-2019-2-57-67

Open challenges in visual analytics for security information and event management

E. S. Novikova[a,b], PhD, Tech, Assosiate Professor, orcid.org/0000-0003-2923-4954, novikova.evgenia123@gmail.com
I. V. Kotenko[a], Dr. Sc., Tech, Professor, orcid.org/0000-0001-6859-7120
[a]Saint-Petersburg Institute for Informatics and Automation of the RAS, 39, 14 Line, V. O., 199178, Saint-Petersburg, Russian Federation
[b]Saint-Petersburg Electrotechnical University «LETI», 5, Prof. Popov St., 197376, Saint-Petersburg, Russian Federation

Introduction: Visual analytics techniques support efficient analysis of the ever-growing amounts of data generated by security sensors and facilitate a timely and reasonable response to the threats. Modern security information and event management systems propose various solutions for processing large data streams and integrating heterogeneous sources which can be used as a framework

to construct a visual analytics system for security tasks. **Purpose:** The analysis of visual analytics techniques implemented in security information and event management systems and designed to support the studies on security incidents in the context of the main visual analytics problems, including the validation of automatic analysis models. **Results:** A contradiction has been detected between the capabilities of security information and event management systems in the visual analysis of security data and the implementation of these capabilities. Techniques for visual correlation of the data from different security sensors and for visual validation of automatic analysis models which would allow you to evaluate their accuracy and adaptability to the changes in data streams are almost missing. A possible way to resolve this contradiction is using techniques which support a flexible mechanism for adjusting the analyzed attributes of the network device events. The article presents the main approaches to the development of such techniques, discussing their advantages and disadvantages. We propose a dashboard for monitoring the behavior of an automated network traffic analysis model used in a cloud computing infrastructure. It allows you to monitor the analysis model behavior, perform a visual correlation of the analyzed parameters, and track changes in the network flows. **Practical relevance:** The results of the research can be used when designing security visual analytics tools for monitoring data flows and the behavior of automated analysis models.

Keywords — visual analytics, security information and event management, visual data correlation, visual validation of analysis models, matrix-base data visualization.

For citation: Novikova E. S., Kotenko I. V. Open challenges in visual analytics for security information and event management. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2019, no. 2, pp. 57–67 (In Russian). doi:10.31799/1684-8853-2019-2-57-67

References

1. Marty R. *Visual Analytics — Delivering Actionable Security Intelligence*. Available at: https://www.blackhat.com/us-16/training/visual-analytics-delivering-actionable-security-intelligence.html (accessed 16 January 2019).
2. Guimarães V. T., Freitas C. M. D. S., Sadre R., Tarouco L. M. R. and Granville L. Z. A survey on information visualization for network and service management. *IEEE Communications Surveys & Tutorials*, 2016, vol. 18, no. 1, pp. 285–323. doi:10.1109/COMST.2015.2450538
3. Best D. M., Endert A., Kidwell D. 7 key challenges for visualization in cyber network defense. *Proc. of the Eleventh Workshop on Visualization for Cyber Security (VizSec 2014)*, ACM, 2014, pp. 33–40. doi:10.1145/2671491.2671497
4. Stange J.-E., Dörk M., Landstorfer J., and Wettach R. Visual filter: graphical exploration of network security log files. *Proc. of the Eleventh Workshop on Visualization for Cyber Security (VizSec 2014)*, ACM, 2014, pp. 41–48. doi:10.1145/2671491.2671503
5. 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).
6. Zhang T., Wang X., Li Z., et al. A survey of network anomaly visualization. *Sci. China Inf. Sci.*, 2017, vol. 60, pp. 121101. doi:10.1007/s11432-016-0428-2
7. Krause J., Perer A., Bertini E. Using visual analytics to interpret predictive machine learning models. *Proc. of the 2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016)*, New York, NY, USA. 2016. Available at: https://arxiv.org/abs/1606.05685 (accessed 16 January 2019).
8. Kotenko D. I., Kotenko I. V., Saenko I. B. Methods and tools for attack modeling in large computer networks: state of the problem. *Trudy SPIIRAN* [SPIIRAS Proceedings], 2012, no. 3 (22), pp. 5–30 (In Russian).
9. Kotenko I. V., Stepashkin M. V. Security analysis of computer networks based on modeling by malefactors actions and constructing attack graph. *Trudy Instituta sistemnogo analiza Rossijskoj akademii nauk* [Proceeding of the Institute for Systems Analysis of the Russian Academy of Science], 2007, vol. 31, pp. 126–207 (In Russian).
10. Kotenko I. V., Saenko I. B. Architecture of the system of intelligent information security services in critical infrastructures. *Trudy SPIIRAN* [SPIIRAS Proceedings], 2013, no. 1 (24), pp. 21–40 (In Russian).
11. Few S. *Information dashboard design the effective visual communication of data*. New York, O'Reilly, 2006. 224 p.
12. *OSSIM SIEM system*. Available at: https://www.alienvault.com/products/ossim (accessed 16 January 2019).
13. *Qradar Pulse Visual analytics application of the IBM Radar SIEM-system*. Available at: https://www.ibm.com/support/knowledgecenter/en/SS42VS_7.2.8/com.ibm.apps.doc/c_Qapps_Pulse_intro.html (accessed 16 January 2019).
14. *Splunk Enterprise SIEM system*. Available at: https://www.splunk.com/ru_ru/products/splunk-enterprise.html (accessed 16 January 2019).
15. *Gartner Magic Quadrant for Security Information and Event Management*. Available at: https://www.gartner.com/doc/3894573 (accessed 16 January 2019).
16. *Interactive Discovery component for the ArcSight SIEM-system*. Available at: https://community.softwaregrp.com/t5/ArcSight-Interactive-Discovery/tkb-p/arcsight-interactive-discovery (accessed 16 January 2019).
17. Walton S., Maguire E., and Chen M. Multiple queries with conditional attributes (QCATs) for anomaly detection and visualization. *Proc. of the Eleventh Workshop on Visualization for Cyber Security (VizSec 2014)*, ACM, 2014. doi:10.1145/2671491.2671502
18. Humphries C., Prigent N., Bidan C., and Majorczyk F. CORGI: combination, organization and reconstruction through graphical interactions. *Proc. of the Eleventh Workshop on Visualization for Cyber Security ((VizSec 2014)*, ACM, 2014, pp. 57–64. doi:10.1145/2671491.2671494
19. Landstorfer J., Herrmann I., Stange J., Dork M., and Wettach R. Weaving a carpet from log entries: A network security visualization built with cocreation. *Proc. of the Visual Analytics Science and Technology (VAST)*, Paris, 2014, pp. 73–82. doi:10.1109/VAST.2014.7042483
20. Fischer F., Keim D. A. NstreamAware: real-time visual analytics for data streams to enhance situational awareness. *Proc. of the Eleventh Workshop on Visualization for Cyber Security (VizSec 2014)*, ACM, 2014, pp. 65–72. doi:10.1145/2671491.2671495
21. Cappers B. C. M., van Wijk J. SNAPS: Semantic network traffic analysis through projection and selection. *Proc. of the IEEE Symposium on Visualization for Cyber Security (VizSec 2015)*, 2015, pp. 1–8. doi:10.1109/ VIZSEC.2015.7312768
22. Stone M. In Color Perception, Size Matters. *IEEE Computer Graphics and Applications*, 2012, vol. 32(2), pp. 8–13. doi:10.1109/MCG.2012.37
23. Borisenko K., Smirnov A., Novikova E., and Shorov A. DDoS attacks detection in cloud computing using data mining techniques. *Proc. of the Industrial Conference on Data Mining (ICDM'2016)*, LNAI, vol. 9728, pp. 1–15. doi:10.1007/978-3-319-41561-1_15
24. McKenna S., Staheli D., Fulcher C., Meyer M. BubbleNet: A cyber security dashboard for visualizing patterns. *Computer Graphics Forum*, 2016, vol. 35, pp. 281–290. doi:10.1111/cgf.12904