УДК 004.056.53

А.А. Проноза, А.А. Чечулин, И.В. Котенко

МАТЕМАТИЧЕСКИЕ МОДЕЛИ ВИЗУАЛИЗАЦИИ В SIEM-СИСТЕМАХ

Проноза А.А., Чечулин А.А., Котенко И.В. **Математические модели визуализации в SIEM-системах.**

Аннотация. В статье предложены математические модели визуализации данных в SIEM-системах. Модели визуализации служат для формализации трех основных этапов процесса визуализации. На первом этапе предлагаются модели, с помощью которых происходит унификация сведений об объектах компьютерной сети, имеющих разнородные структуры и различные источники. На втором этапе на базе построенных моделей формируется многомерная матрица связей. На третьем этапе предлагается унифицированный подход к визуализации различных аспектов безопасности компьютерной сети на основе построенной матрицы.

Ключевые слова: визуализация данных безопасности, системы управления событиями и информацией безопасности (SIEM), математические модели визуализации.

Pronoza A.A., Chechulin A.A., Kotenko I.V. **Mathematical Models of Visualization in SIEM Systems.**

Abstract. The paper suggests the mathematical models of data visualization in SIEM-systems. The visualization models formalize three main stages of the visualization process. At the first stage the models are being suggested which fulfill the unification of data on the computer network objects having heterogeneous structures and different sources. At the second stage, on the basis of the suggested models, a multidimensional matrix of relations is generated. At the third stage a uniform approach to the visualization of various security aspects of the computer network on the basis of constructed matrix is proposed.

Keywords: security data visualization; security information and event management (SIEM) systems; mathematical models of visualization.

1. Введение. В настоящее время в распределенных компьютерных сетях происходит стремительный рост количества хостов сети и связей между ними. Каждый хост или их связь генерирует множество сообщений, связанных с безопасностью. Автоматизированный анализ и визуализация указанных сообщений критически важны для предотвращения угроз компьютерной безопасности.

Комплексная обработка сообщений безопасности, поступающих от всех узлов компьютерной сети, происходит в рамках систем управления событиями и информацией безопасности (Security Information and Event Management — SIEM). Подобные системы способны не только эффективно выявлять угрозы безопасности, но и моделировать возможные сценарии компьютерных атак [1]. Важным компонентом SIEM-систем является компонент визуализации. Этот компонент позволяет оператору оперативно и в понятном для него виде получать информацию о состоянии безопасности компьютерной сети, ее уязви-

мых участках и возможных действиях нарушителя [2–5]. Значимость подсистемы визуализации сложно переоценить, поскольку на основе отображаемых данных пользователь системы оценивает текущую обстановку и выбирает соответствующие контрмеры. Неудачная визуализация может привести к снижению качества выбираемых контрмер и, как следствие, к понижению уровня защищенности компьютерной сети.

К основным графическим моделям, используемым при визуализации информации, можно отнести графики, графы и их вариации, матрицы, а также карты [8]. При использовании классических моделей для наглядного представления данных применяются некоторые визуальные приемы, такие как расположение объекта в пространстве, кодирование информации при помощи его формы, размера и цвета [7]. Формируемое изображение должно быть понятным и информативным и не должно выходить за рамки возможностей когнитивного аппарата человека [8].

Визуализация может быть дополнена рядом инструментов, таких как «рыбий глаз» [8], «семантическое масштабирование» [10], «небольшие различия» [11] и другими. Важно также предоставить пользователю возможность получить исчерпывающую информацию о выбранном объекте при помощи визуального средства поиска [12].

Классическим представлением модели компьютерной сети является граф, в котором под узлами понимаются хосты сети, а дугами обозначаются связи между хостами. В работах [16, 17] рассмотрен подход, в котором для изображения узлов графа предлагается отображать метрики защищенности соответствующих хостов. Указанный подход расширен в работе [18], в результате чего у пользователя появляется возможность видеть как текущие, так и предыдущие значения метрик безопасности.

При исследовании вопросов визуализации больших компьютерных сетей основное внимание уделяется вопросам, связанным с отображением информации в условиях ограниченной поверхности. Например, в работе [13] предлагается способ визуализации сложной сети, состоящей из реальных объектов и связей между ними, но понятия «объект» и «связь» не раскрываются в объеме, необходимом для их графического представления. В работе [14] представлены модели визуализации маршрутов атак в виде графов и плоских карт деревьев, однако отсутствует описание связи между компьютерной инфраструктурой и ее графическим представлением. В работах, посвященных визуализации, предполагается, что сведения об объектах компьютерной инфраструктуры и ее структуре заранее известны и подготовлены для отображения.

Однако сам способ представления данных может существенно повлиять на результат визуализации. Например, в [15] описывается

процесс построения графов атак и расчет метрик защищенности, основанный на формировании для каждого хоста трехмерной матрицы по следующим данным: (1) класс атак; (2) необходимый тип доступа; (3) уровень знаний нарушителя. Такой подход позволяет строить многоуровневые модели сценариев атак с их последующей визуализацией при помощи классических моделей, таких как графы.

Особенности представления и интерпретации одних и тех же данных могут быть использованы для взаимодействия с пользователем в рамках одной модели визуализации, не исключая использование стандартных инструментов. Такое представление данных можно назвать интерактивным, поскольку оно позволяет пользователю посмотреть на одну и ту же систему под разными углами.

С точки зрения авторов, подсистема визуализации SIEM-системы не только обеспечивает визуализацию сообщений безопасности, но и является аналитической системой обработки информации. Каждый запрос пользователя на визуализацию каких-либо параметров компьютерной сети заставляет подсистему визуализации проводить соответствующую выборку из огромного массива разнородных данных, сохраняемых SIEM-системой.

Процессу отображения информации должен предшествовать процесс ее обработки и приведение к виду, удобному для отображения, при этом достигается его интерактивность. Таким образом, представляется необходимым с теоретической точки зрения рассмотреть структуры данных, использующиеся при визуализации компьютерной инфраструктуры, и способ их организации для единообразного получения различных представлений компьютерной сети.

Целью настоящей работы является построение математических моделей представления разнородных данных о компьютерной сети и демонстрация того, как построенные модели могут быть визуализированы при помощи древовидных и графовых структур.

Как будет показано ниже, организация данных, собираемых SIEM-системой, носит многомерный характер. На рисунке 1 представлена предлагаемая схема функционирования подсистемы визуализации, основанная на предложенном в [19] общем подходе к анализу многомерных данных.

В качестве исходных данных мы рассматриваем поступающую в SIEM-систему информацию о компьютерной сети, которая используется для заполнения атрибутов моделей информационных объектов — компьютерной сети, хостов, и связей между хостами. В силу разнородности этой информации возникает необходимость в ее формализа-

ции и унификации, чтобы в дальнейшем получить визуализацию состояния безопасности всей сети.

Рис. 1. Общая схема анализа данных большой размерности

Предложенные в настоящей статье математические модели информационных объектов описывают наиболее важные компоненты компьютерной инфраструктуры с учетом как их формальных признаков (например, установленное на хостах программное обеспечение и наличие в нем уязвимостей), так и логических (например, метрики безопасности). В рамках общего подхода к визуализации компьютерной сети указанные модели объединяются в многомерную матрицу связей, в которой наличие связи между двумя моделями определяется типом отношения, в котором состоят исходные информационные объекты. После чего на базе этой матрицы строятся универсальные модели ее графического представления.

Особенностью предложенных в статье моделей представления данных является единообразный подход к их организации, при котором учитываются имеющиеся различия в их природе и структуре. Указанный подход позволяет использовать единую модель визуализации компьютерной сети, которая, в зависимости от параметра, способна отображать разные аспекты безопасности этой сети.

2. Математические модели информационных объектов. Рассмотрим компьютерную сеть N, состоящую из множества хостов H и связей между ними C:

$$N = (H, C). \quad (1)$$

Под информационным объектом $o \in O$ будем подразумевать хост $h \in H$, связь $c \in C$ или подсеть $N' \subseteq N$.

Модель хоста h может включать в себя огромное количество различных атрибутов, от значения которых зависит состояние безопасности хоста. В настоящей статье приводится набор атрибутов, наиболее важных с точки зрения авторов, без которых невозможно определить, каким образом характеристики хоста влияют на безопасность всей сети.

Построим модель хоста следующим образом.

$$h = (ID, P_{hard}, P_{soft}, V, t^h, I_{cr}, Sc), \quad (2)$$

где ID — идентификационная информация, $P_{hard} \subset P$ — установленное аппаратное обеспечение, $P_{soft} \subset P$ — установленное программное обеспечение, V — множество уязвимостей, $t^h \in T^h$ — тип сетевого объекта, I_{cr} — важность обрабатываемой информации, Sc — совокупность характеристик, которые можно собрать о хосте в результате внешнего сканирования.

Множество P содержит сведения о программном и аппаратном обеспечении, такие как фирма-производитель, название программного продукта и его версия; множество T^h содержит все возможные типы хостов (автоматизированное рабочее место, сервер, маршрутизатор и т.п.).

Атрибуты модели могут задаваться как в автоматическом режиме (например, множество уязвимостей может быть получено в результате работы сканера уязвимостей), так и в ручном режиме (например, администратор безопасности самостоятельно определяет важность обрабатываемой на хосте информации).

Модель связи между хостами должна включать в себя непосредственно хосты h_i и h_j, физические каналы передачи данных T', существующие между хостами, а также аппаратные и программные средства U', установленные на хостах и способные вести информационный обмен. Таким образом, модель связи c примет вид:

$$c = \left(h_i, h_j, T', U'\right), h_i, h_j \in H, T' \subset T, U' \subset U. \quad (3)$$

Модели каналов передачи данных включают в себя такие обязательные атрибуты, как тип канала, протокол передачи, а также ха-

рактеристики источника и приемника, без которых невозможна передача данных.

Приведенные в (1), (2) и (3) модели описывают основные объекты компьютерной инфраструктуры, информацию о которых должна собирать и обрабатывать SIEM-система.

Состояние информационных объектов определяется различными источниками информации, такими как программно-аппаратные комплексы мониторинга безопасности, средства анализа сети, коммутационное оборудование, серверы обновлений программного обеспечения и т.п.

Источники информации можно разделить на физические и логические.

Под физическим источником информации следует понимать произвольный узел сети N, в том числе и внешний по отношению к этой сети, на котором установлено программное обеспечение, способное отслеживать и протоколировать события безопасности. В общем случае математическая модель формализованного события безопасности будет иметь вид:

$$a = \{(type, source, severity, timestamp, msg)\}, a \in A, \qquad (4)$$

где $type$ — тип сообщения, $source$ — источник, сгенерировавший сообщение, $severity$ — важность сообщения, $timestamp$ — временная метка. Тело сообщения msg представляет собой структуру, поля которой содержат специфические для информационного объекта или его компонента характеристики.

Под логическим источником будем понимать компонент системы оценки защищенности информационного объекта, т.е. узел сети N, способный вычислять количественные показатели безопасности информационных объектов, такие как уровень критичности, степень уязвимости, вероятность успешной атаки и т.п. [18]. Для набора совокупностей информационных объектов $O = \{O_1, .., O_k\}$ математическая модель метрики безопасности f примет вид:

$$f(\mathrm{O}) \to \mathbb{R}. \qquad (5)$$

Таким образом, на основании введенной классификации источников информации, математическая модель сбора информации со всех физических источников L_p и всех логических источников L_l описывается следующим образом:

$$\mathcal{M} = (Q, W), \qquad (6)$$

где образ отображения $Q: L_p \to A$ представляет собой набор формализованных сообщений безопасности, а образ отображения $W: L_l \to B$

является набором числовых показателей защищенности информационных объектов.

Полученная таким образом информация должна быть передана модулю сбора данных подсистемы визуализации для последующей обработки, анализа и визуализации. Способ передачи зависит от источника информации, однако, как правило, внутренние источники самостоятельно инициируют передачу данных, внешние источники опрашиваются с заданным временным интервалом.

3. Многомерная матрица связей информационных объектов. Существующие в компьютерной инфраструктуре информационные объекты, такие как хосты и связи между ними, могут быть объединены в единую модель, представленную в виде многомерной матрицы связей. Как будет показано ниже, в результате применения операций среза к указанной матрице могут быть получены различные представления компьютерной сети, акцентирующие внимание администратора безопасности на конкретных проблемных моментах сети.

Для построения многомерной матрицы связи введем следующую классификацию отношений T^c между хостами:

– отношение физической доступности, при котором между двумя хостами существует физический канал связи;

– отношение доступности, при котором имеется возможность установить канал передачи данных с помощью сетевого протокола;

– отношение посредством виртуальных каналов связи, при котором между хостами существуют логические каналы связи, обладающие специфическими с точки зрения безопасности характеристиками, например, шифрованием трафика;

– отношение функциональной зависимости, при котором для связи с главным хостом необходимо наличие связи с зависимым хостом.

– отношение доверия, при котором существующий между хостами канал связи не контролируется политиками безопасности;

– отношения уязвимости, при котором эксплуатация уязвимости на одном хосте ведет к компрометации другого хоста.

Введем многомерную матрицу связей, как:

$$M^c = \left\| m_{ijk} \right\|, i, j = 1..|H|, k = 1..|T^c|, \tag{7}$$

ее элементы получаются в результате отображения:

$$F^c: (\mathbb{N}, \mathbb{N}, \mathbb{N}) \to H \times H \times T^c,$$

$$m_{ijk} = F^c(i, j, k) = \begin{cases} (h_i, h_j, t_k, msg), \text{если } h_i, h_j \text{ в отношении } t_k \\ (h_i, h_j, 0), \text{в противном случае} \end{cases}. \tag{8}$$

Таким образом, элементами матрицы M^c являются наборы (h_i, h_j, t_k, msg), представляющие собой упорядоченную пару хостов и тип связи (отсутствие связи) между ними. Значение атрибута msg может содержать дополнительную информацию по конкретному типу связи.

Введем процедуру формирования среза матрицы M^c через фиксацию индекса $k = \tilde{k}$:

$$\pi(\tilde{k}) = \left\| m_{ij\tilde{k}} \right\|, i, j = 1..|H|. \quad (9)$$

Отметим, что каждый срез $\pi(\tilde{k})$ представляет собой матрицу смежности графа, вершинами которого являются хосты, а наличие ребра определяет наличие между хостами соответствующей связи. Интерпретация каждого среза приведена в таблице 1.

Таблица 1. Результаты функции π(k) и их интерпретация

$\pi(k)$	Интерпретация
$\pi(1)$	Топология сети
$\pi(2)$	Граф доступности
$\pi(3)$	Граф защищенных каналов
$\pi(4)$	Граф зависимостей
$\pi(5)$	Граф доверия
$\pi(6)$	Граф атак

При визуализации многомерной матрицы процедура формирования среза матрицы $\pi(k)$ решает проблему восприятия пользователем многомерных данных, заключающуюся в том, что человек способен воспринимать изображения размерностью $n \leq 3$.

4. Математическая модель визуализации многомерной матрицы связей. Для визуализации компьютерных сетей в целом, и характеристик их безопасности в частности, обычно используются древовидные и графовые структуры. Однако построение математических моделей визуализации указанных структур обладает рядом особенностей, возникающих в силу ограниченного размера поверхности, на которую происходит их отображение.

Кроме исходных данных, таких как срез многомерной матрицы связей $\pi(\tilde{k})$, следует также учитывать необходимость масштабирования получаемого изображения и его укрупнения, т.е. агрегацию отдельных узлов сети с целью отобразить наиболее общую картину (семантическое масштабирование [10]). С другой стороны, для уменьшения числа связей зачастую вместо графа целесообразнее отображать соответствующее остовное дерево.

В общем случае срез матрицы $\pi(\tilde{k})$ отображается в виде неориентированного графа $G = (V, E)$, где V — вершины графа, представляющие собой хосты, E — ребра графа, задающие связи между хостами.

Определим математическую модель визуализации указанного графа следующим образом:

$$\chi_{\mathrm{k}}(\pi(k), V_{scale}, G_{type}) = \mathcal{G}(G'_k), k = 1..|T^c|,$$
$$G'_k = (V_{scale}, E'), V_{scale} \subseteq V, E' \subseteq E, \quad (10)$$

где $\pi(k)$ — соответствующий срез матрицы M^c, G'_k — подграф G_i, содержащий вершины из множества V_{scale}, $\mathcal{G}$ — процедура визуализации графа. Множества V_{scale} и G_{type} задаются пользователем и обеспечивают геометрическое масштабирование и компоновку визуального представления соответственно. Процедура визуализации графа $\mathcal{G}$ обеспечивает вывод графа на экран способом, указанным в G_{type}.

Выделение маршрутов в графе может быть представлено при помощи изменения цветовой схемы ребер из соответствующего набора H':

$$\eta^e(H') = color(\{e_{ij} | e_{ij} = (v_i, v_j), v_i, v_j \in H'\}). \quad (11)$$

Процесс агрегации узлов сети стоит в том, чтобы объединять узлы графа по определенному критерию, например, по их принадлежности к одной подсети или домену. Результатом агрегации узлов будет набор подграфов G_i, соединенных между собой хотя бы одним ребром $e \in E$:

$$G = \{G_i | G_i \subset G, i = 1..p\}. \quad (12)$$

Уменьшение числа связей заключается в построении остовного дерева графа G. Например, при отображении топологии сети «звезда» вершина с максимальной степенью является основным связующим звеном сети. На практике такой вершиной может быть сетевой коммутатор или точка доступа Wi-Fi.

Рассмотрим функцию θ, формирующую остовное дерево для графа G:

$$T = \theta(G) = (V, E'), E' \subseteq E. \quad (13)$$

Построим для каждого подграфа G_i его остовное дерево:

$$T_i = \theta(G_i) = (V_i, E_i'), E_i{}' \subseteq E_i. \quad (14)$$

Совокупность всех остовных деревьев $\{T_i\}_{i=1}^{p}$, соединенных между собой, является иерархической структурой сети, узлы которой представляют собой подсети.

Рассмотренный подход предполагает двухуровневую кластеризацию сетей, однако может быть расширен для подсетей любого уровня.

Комбинированное применение описанных выше подходов обеспечивает читаемое представление графа на ограниченной поверхности с настраиваемым количеством отображаемых деталей.

Рассмотрим два подхода к визуализации отдельного хоста — узла графа G. Выбор конкретного подхода зависит от решаемых задач и осуществляется пользователем.

Первый подход заключается в сопоставлении каждому хосту его пиктограммы, содержащей информацию, например, о типе узла. В этом случае модель визуализации хоста представляется функцией:

$$D^h(t^h) = image^h, \tag{15}$$

где t^h — тип хоста, $image^h$ — соответствующая типу хоста пиктограмма.

Второй подход заключается в сопоставлении каждому хосту глифа, т.е. изображения, на котором отображены метрики безопасности, подсчитанные для данного хоста. Глиф представляет собой масштабируемое изображение, разделенное на секции. Каждая секция соответствует определенной метрике безопасности и, в зависимости от критичности значения этой метрики, может выделяться соответствующим цветом.

Для того чтобы сделать модель глифа более информативной, в нее следует включить историю изменений значений метрик безопасности, располагающуюся по контуру соответствующей секции и обозначенной цветом, соответствующим предыдущему значению этой метрики [18].

Базой глифа B_t^h для хоста h в момент времени t назовем множество вида:

$$B_t^h = \left\{ f_i \middle| f_i\left(O_i^h\right) = x_i \right\}, i = 1..n, x_i \in \mathbb{R}, \tag{16}$$

где $O_i^h \subseteq O$ — набор совокупностей информационных объектов, содержащий сведения о хосте h, на котором имеет смысл отображение соответствующей метрики безопасности $f_i \in F$. В частности, такой набор может состоять из одного единственного хоста h.

Математическая модель глифа имеет вид:

$$G_k^h = \left\{ B_{t_1}^h, \dots, B_{t_k}^h \right\}, k \geq 1. \tag{17}$$

Построенная таким образом модель глифа позволит администратору проводить анализ защищенности объектов сетевой инфраструктуры во времени и наблюдать изменения состояния безопасности хоста в зависимости от принятых мер по администрированию сети.

5. Пример применения моделей. Рассмотрим применение предлагаемых моделей для визуализации топологии сети и графа атак на этапах сбора и обработки информации, получаемой от физических и логических источников, построения многомерной матрицы связей и ее визуализации. Отметим, что количество сведений о компьютерной инфраструктуре, собираемое подсистемой визуализации, для отдельного вида отношений многомерной матрицы связей будет избыточно.

Для визуализации топологии сети на этапе сбора информации наиболее важные сведения поступают из сообщений физических источников. Из всего множества сообщений A необходимо выделить только те из них, которые содержат информацию о хостах и физическом канале связи между ними:

$$A_p = \{a | a \in I(h_i, h_j, T'), a \in A\}. \quad (18)$$

Множество сообщений A_p позволяет заполнить атрибуты модели связи (3), при этом можно принять $U' = \emptyset$. Сформированные модели, в свою очередь, используются для формирования многомерной матрицы связей (8) при $k = 1$ (отношение физической доступности). Результатом среза матрицы $\pi(1)$ будет матрица смежности, задающая неориентированный граф, вершинами которого являются хосты, а ребрами — физические каналы связи. Следует отметить, что в случае сетевой топологии «звезда» при визуализации (10) полученного графа можно разбить компьютерную сеть на сегменты, принимая вершины с максимальной степенью за коммутаторы, как показано на рисунке 2.

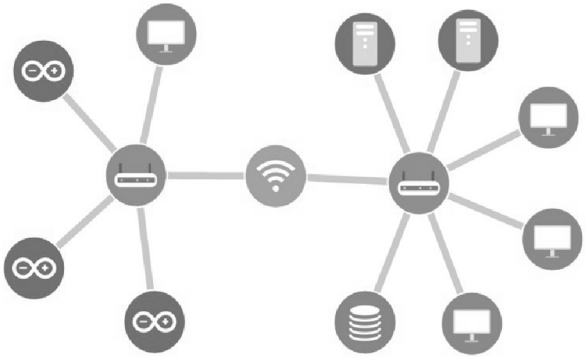

Рис. 2. Пример визуализации сети, состоящей из двух сегментов

Модель формирования и визуализации графа атак отличается от описанной выше модели топологии сети, однако может быть построена по тому же алгоритму.

Источниками сообщений для формирования графа атак являются как физические, так и логические источники. Из множества сообщений A от физических источников необходимо выделить только те из них, которые содержат информацию о хостах, их уязвимостях и установленном программно-аппаратном обеспечении. При этом из указанного множества следует исключить хосты без уязвимостей, а для оставшихся хостов выбрать наиболее критичную уязвимость:

$$A_a = \{a | a \in I(h_i, v) \vee a \in I(h_i, p), severity(v) \geq \max\{severity(V_{h_i})\}, h_i \in v, v \in V, p \in P, a \in A\}. \quad (19)$$

Из множества сообщений B от логических источников следует выбрать сообщения, позволяющие оценить степень уязвимости V_{level} хоста:

$$B_a = \{f(0) | f(0) \in V_{level}\}. \quad (20)$$

В общем случае граф атак является подграфом графа доступности, описываемого при помощи среза $\pi(5)$ многомерной матрицы связей. Граф атак, также как и граф доступности, является ориентированным, и направление его ребер задается исходя из атрибутов моделей связи между хостами (3).

В некоторых случаях, например, при рассмотрении DoS-атак, граф атак будет являться подграфом графа зависимостей.

Также следует отметить, что в результате обработки входных данных может получиться несколько графов атак, каждый из которых будет подграфом графа доступности. В этом случае их следует рассматривать по отдельности.

Особенностью визуализации графа атак является, с одной стороны, выделение на графе доступности возможных направлений атак, исходя из анализа уязвимостей, расположенных на связанных хостах (направления возможных атак представлено в виде стрелок); с другой стороны — использование глифов вместо пиктограмм хостов. Использование глифов позволяет визуализировать 4 характеристики хоста (каждая имеет три возможных варианта значения, отображаемых синим, желтым и зеленым цветами) и их изменение со временем (цвет внутреннего сектора и внешней линии). В качестве отображаемой характеристики может выступать важность информации (I_{cr}), наличие уязвимостей разного уровня критичности (V) и т.д. Пример графа атак приведен на рисунке 3.

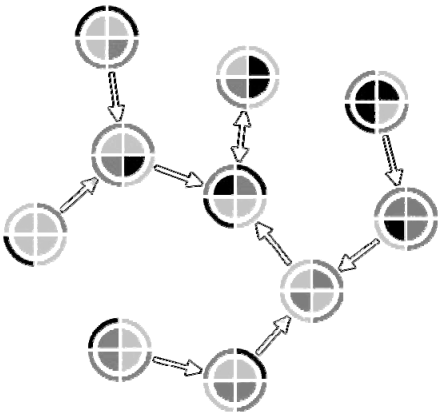

Рис. 3. Пример графа атак с использованием глифов

Таким образом, построенные модели описывают процесс визуализации состояния компьютерной инфраструктуры в SIEM-системах с целью акцентировать внимание пользователя на различных аспектах безопасности сети при помощи единой системы сбора, обработки, представления и отображения информации.

6. Заключение. Предлагаемый в настоящей статье подход призван компенсировать недостаток информации о процессе сбора и структурирования данных разнородной структуры о компьютерной сети для SIEM-систем. Данный процесс проходит внутри подсистемы визуализации и во многом предшествует процессу отображения информации. Для его описания предлагаются математические модели визуализации компьютерной сети с использованием графовых и древовидных структур.

В статье введено понятие информационного объекта как обязательной составляющей компьютерной инфраструктуры, и представлены математические объекты, содержащие необходимые для анализа безопасности компьютерной сети атрибуты.

На основании классификации типов отношений между информационными объектами построена многомерная матрица связей информационных объектов между собой и показано, как срезы указанной матрицы могут представлять компьютерную инфраструктуру под разными углами и акцентировать внимание оператора SIEM-системы на конкретных проблемных участках сети.

Далее представлены математические модели визуализации срезов многомерной матрицы связей в виде графа, в которых учтены вопросы сематического масштабирования (агрегирования узлов) и

уменьшения количества связей между узлами. Описывается возможность использования глифов для отображения метрик безопасности как отдельных хостов сети, так и ее сегментов (узлов графа) [20]. Также в статье представлены элементы интерфейса разрабатываемого в настоящее время программного прототипа компонента визуализации информации безопасности на основе веб-интерфейса. Предполагается, что в дальнейшем этот компонент может быть использован в существующих SIEM-системах.

Построение математических моделей для визуализации позволяет выделить наиболее значимые аспекты исследуемой области, описать их структуру в виде набора необходимых для исследования атрибутов, скомпоновать разнородные массивы данных в единую многомерную структуру и представить процесс визуализации как отображение ее основных срезов с использованием различных графических моделей. Эта задача очень важна для SIEM-систем, так как эффективность принятия решений оператором напрямую зависит от оперативности и адекватности восприятия им информации, выдаваемой SIEM-системой.

В дальнейших исследованиях планируется разработать математические модели для визуализации различных данных, хранящихся в реляционных и онтологических хранилищах SIEM-систем [21] на основе таких моделей визуализации, как графы, глифы, карты деревьев и диаграмма Вороного. Также планируется разработать отдельный компонент визуализации, основанный на веб-формах, который мог бы быть включен в большинство существующих SIEM-систем.

Литература

1. *Miller D., Harris S., Harper S., VanDyke C.* Security Information and Event Management (SIEM) Implementation // McGraw Hill Professional. 2010. 464 p.
2. Официальный сайт системы Arcsight SIEM. URL: http://www.ndm.net/siem/main/arcsight-siem (дата обращения 15.02.16).
3. Официальный сайт системы NetIQ Sentinel SIEM. URL: https://www.netiq.com/products/sentinel/ (дата обращения 15.02.16).
4. Официальный сайт системы QRadar SIEM. URL: http://q1labs.com/products/qradar-siem.aspx (дата обращения 15.02.16).
5. Официальный сайт системы OSSIM SIEM. URL: http://communities.alienvault.com/ (дата обращения 15.02.16).
6. *Коломеец М.В., Чечулин А.А., Котенко И.В.* Обзор методологических примитивов для поэтапного построения модели визуализации данных // Труды СПИИРАН. 2015. Вып. 42. С. 232–257.
7. *Клышинский Э.С., Рысаков С.В., Шихов А.И.* Обзор методов визуализации многомерных данных // Новые информационное технологии в автоматизированных системах. 2014. С. 519–530.
8. *Goldstein B.* Cognitive Psychology // Wadsworth, Cengage Learning. 2011. 472 p.
9. *Sarkar M., Brown M.* Graphical fisheye views // Communications of the ACM. 1994. vol. 37. no. 12. pp. 73–83.

10. *Watson G.* Lecture 15 – Visualisation of Abstract Information // Edinburgh Virtual Environment Centre. 2004.
11. *Wroblewski L.* Small Multiples within a User Interface // Web Form Design. 2005.
12. *Ferebee D., Dasgupta D.* Security Visualization Survey // Proceedings of the 12th Colloquium for Information Systems Security Education University of Texas. 2008. 124 p.
13. *Пупырев С.Н.* Модели, алгоритмы и программный комплекс визуализации сложных сетей // Автореферат диссертации на соискание ученой степени кандидата физико-математических наук. Екатеринбург. 2010. 21 с.
14. *Котенко И.В., Новикова Е.С.* Визуальный анализ для оценки защищенности компьютерных сетей // Информационно-управляющие системы. 2013. № 3. С. 55–61.
15. *Kotenko I.V., Doynikova E.V., Chechulin A.A.* Security metrics based on attack graphs for the Olympic Games scenario // IEEE Computer Society. 2014. pp. 561–568.
16. *Erbacher R.* Visualization Design for Immediate High-Level Situational Assessment // Proceedings of International Symposium on Visualization for Cyber Security (VizSec'12). 2012. pp.17–24.
17. *Matuszak W.J., DiPippo L., Lindsay Y.* Sun CyberSAVe – Situational Awareness Visualization for Cyber Security of Smart Grid Systems // Proceedings of International Symposium on Visualization for Cyber Security (VisSec'13). 2013. pp. 25–32.
18. *Kotenko I.V., Novikova E.S.* Visualization of Security Metrics for Cyber Situation Awareness // IEEE Computer Society. 2014. pp. 506–513.
19. *Бондарев А.Е., Галактионов В.А.* Анализ многомерных данных в задачах многопараметрической оптимизации с применением методов визуализации // Научная визуализация. 2012. Вып. 4. № 2. С. 1–13.
20. *Котенко И.В, Степашкин М.В., Дойникова Е.В.* Анализ защищенности автоматизированных систем с учетом социо-инженерных атак // Проблемы информационной безопасности. Компьютерные системы. 2011. № 3. С. 40–57.
21. *Kotenko I., Polubelova O., Saenko I.* The Ontological Approach for SIEM Data Repository Implementation // IEEE Computer Society. 2012. pp. 761–766.

References

1. Miller D., Harris S., Harper S., VanDyke C. Security Information and Event Management (SIEM) Implementation. McGraw Hill Professional. 2010. 464 p.
2. Oficial'nyj sajt sistemi Arcsight SIEM [Official web site of Arcsight SIEM]. Available at: http://www.ndm.net/siem/main/arcsight-siem (accessed 15.02.16).
3. Oficial'nyj sajt sistemi NetIQ Sentinel SIEM [Official web site of NetIQ Sentinel SIEM]. Available at: https://www.netiq.com/products/sentinel/ (accessed 15.02.16).
4. Oficial'nyj sajt sistemi QRadar SIEM [Official web site of QRadar SIEM]. Available at: http://q1labs.com/products/qradar-siem.aspx (accessed 15.02.16).
5. Oficial'nyj sajt sistemi OSSIM SIEM [Official web site of OSSIM SIEM]. Available at: http://communities.alienvault.com/ (accessed 15.02.16).
6. Kolomeec M.V., Chechulin A.A., Kotenko I.V. [Review of methodological primitives for the phased construction of data visualization model]. *Trudy SPIIRAN – SPIIRAS proceedings*. 2015. vol. 42. pp. 232–257. (In Russ.).
7. Klyshinskij E.S., Rysakov S.V., Shihov A.I. [Review of the methods of multidimensional data visualization]. *Novye Informacionnoe Tehnologii v Avtomatizirovannyh Sistemah – New information technologies in automated systems*. 2014. pp. 519–530. (In Russ.).
8. Goldstein B. Cognitive Psychology. Wadsworth, Cengage Learning. 2011. 472 p.
9. Sarkar M., Brown M. Graphical fisheye views. Communications of the ACM. 1994. vol. 37. no. 12. pp. 73–83.

10. Watson G. Lecture 15 – Visualisation of Abstract Information. Edin-burgh Virtual Environment Centre. 2004.
11. Wroblewski L. Small Multiples within a User Interface. Web Form Design. 2005.
12. Ferebee D., Dasgupta D. Security Visualization Survey.Proceedings of the 12th Colloquium for Information Systems Security Education University of Texas. 2008. 124 p.
13. Pupyrev S.N. *Modeli algoritmy i programmnyj kompleks vizualizacii slozhnyh setej* [Models, algorithms and software complex for visualization of complex networks. PhD thesis]. Ekaterinburg. 2010. 21 p. (In Russ.).
14. Kotenko I.V., Novikova E.S. [Visual analysis for evaluation of security of computer networks.]. *Informacionno-upravlyaushie systemy – Information and Control Systems*. 2013. vol. 3. pp. 55–61. (In Russ.).
15. Kotenko I.V., Doynikova E.V., Chechulin A.A. Security metrics based on attack graphs for the Olympic Games scenario. IEEE Computer Society. 2014. pp. 561–568.
16. Erbacher R. Visualization Design for Immediate High-Level Situational Assessment. Proceedings of International Symposium on Visualization for Cyber Security (VizSec'12). 2012. pp. 17–24.
17. Matuszak W.J., DiPippo L., Lindsay Y. Sun CyberSAVe – Situational Awareness Visualization for Cyber Security of Smart Grid Systems. Proceedings of International Symposium on Visualization for Cyber Security (VisSec'13). 2013. pp. 25–32.
18. Kotenko I.V., Novikova E.S. Visualization of Security Metrics for Cyber Situation Awareness. IEEE Computer Society. 2014. pp. 506–513.
19. Bondarev A.E., Galaktionov V.A. [Investigation of multidimensional data in multiparameter optimization with visualization techniques usage]. *Nauchnaia vizualizatia – Scientific visualization*. 2012. vol. 4. no. 2. pp. 1–13. (In Russ.).
20. Kotenko I.V., Stepashkin M.V., Dojnikova E.V. Security analysis of information systems taking into account social engineering attacks]. *Problemy informacionnoj bezopasnosti Kompyuternye sistemy – Problems of information security. Computer systems.* 2011. vol. 3. pp. 40–57. (In Russ.).
21. Kotenko I., Polubelova O., Saenko I. The Ontological Approach for SIEM Data Repository Implementation. IEEE Computer Society. 2012. pp. 761–766.

Проноза Антон Александрович — аспирант лаборатории проблем компьютерной безопасности, Федеральное государственное бюджетное учреждение науки Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов: безопасность распределенных систем, визуализация данных. Число научных публикаций — 4. pronoza@gmail.com, http://pronoza.ru; 14-я линия В.О., д. 39, ком. 205, Санкт-Петербург, 199178; р.т.: 89500394868.

Pronoza Anton Aleksandrovich — PhD student of computer security problems laboratory, St. Petersburg Institute for Informatics and Automation of Russian Academy of Sciences (SPIIRAS).. Research interests: distributed system security, security visualization. The number of publications — 4. pronoza@gmail.com, http://pronoza.ru; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: 89500394868.

Чечулин Андрей Алексеевич — к-т техн. наук, старший 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, анализ сетевого трафика, анализ уязвимостей. Число научных публикаций — 150. andreych@bk.ru, http://comsec.spb.ru/ru/staff/chechulin; 14-я линия В.О., д. 39, ком. 205, Санкт-Петербург, 199178; р.т.: +78123287181.

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.

Котенко Игорь Витальевич — д-р техн. наук, профессор, заведующий лабораторией проблем компьютерной безопасности, Федеральное государственное бюджетное учреждение науки Санкт-Петербургского института 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)328–2642, Факс: +7(812)328–4450.

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)328–2642, Fax: +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.

Acknowledgements. This research is supported by RFBR (projects No. 14-07-00697, 14-07-00417, 15-07-07451, 16-37-00338, 16-29-09482), 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.

РЕФЕРАТ

Проноза А.А., Чечулин А.А., Котенко И.В. **Математические модели визуализации в SIEM-системах.**

В условиях постоянного роста количества разнородных данных, которые должны быть проанализированы оператором системы безопасности, весьма актуальна проблема разработки формальных математических моделей визуализации. Это необходимо как для построения новых подходов к визуализации, так и для повышения эффективности существующих.

В статье предложены математические модели визуализации данных в SIEM-системах. Модели визуализации служат для формализации трех основных этапов процесса визуализации. На первом этапе предлагаются модели, с помощью которых происходит унификация сведений об объектах компьютерной сети, имеющих разнородные структуры и различные источники. На втором этапе на базе построенных моделей формируется многомерная матрица связей. На третьем этапе предлагается унифицированный подход к визуализации различных аспектов безопасности компьютерной сети на основе построенной матрицы.

SUMMARY

Pronoza A.A., Chechulin A.A., Kotenko I.V. **Mathematical Models of Visualization in SIEM Systems.**

Development of new formal mathematical visualization models is an important task, especially when the amount of data, which should be analyzed by the operator, is constantly increasing. This is important both for developing new visualization approaches and enhancing existing ones.

The paper suggests the mathematical models of data visualization in SIEM-systems. The visualization models formalize three main stages of the visualization process. At the first stage the models are being suggested which fulfill the unification of data on the computer network objects having heterogeneous structures and different sources. At the second stage, on the basis of the suggested models, a multidimensional matrix of relations is generated. At the third stage a uniform approach to the visualization of various security aspects of the computer network on the basis of constructed matrix is proposed.