УДК 004.056.53

М.В. Коломеец, А.А. Чечулин, И.В. Котенко

ОБЗОР МЕТОДОЛОГИЧЕСКИХ ПРИМИТИВОВ ДЛЯ ПОЭТАПНОГО ПОСТРОЕНИЯ МОДЕЛИ ВИЗУАЛИЗАЦИИ ДАННЫХ

Коломеец М.В., Чечулин А.А., Котенко И.В., **Обзор методологических примитивов для поэтапного построения модели визуализации данных.**
Аннотация. В статье рассматриваются основные методологические примитивы на примере поэтапного построения модели визуализации с заранее подготовленными данными, с целью сформировать комплексное видение процесса создания модели и влияющих на неё аспектов. Приводится классификация примитивов и их связи между собой в соответствии с этапами построения модели. Рассматриваются библиотеки визуализации на популярных языках программирования.
Ключевые слова: визуализация данных безопасности, обзор методик визуализации, модель визуализации.

Kolomeec M.V., Chechulin A.A., Kotenko I.V. **Review of Methodological Primitives for the Phased Construction of Data Visualization Model.**
Abstract. The paper considers common methodological primitives by the example of phased data visualization model construction with prepared data. The purpose of the article is to form a comprehensive vision of the visualization model creation. The primitives classification and communication between them are suggested. The paper also considers data visualization libraries in popular programming languages.
Keywords: security data visualization, overview, method, visualization model.

1. Введение. Построение модели визуализации это сложный процесс, который определяют различные аспекты [1-2]. При этом, сами аспекты могут принадлежать различным областям знаний, таким как компьютерная графика, математика, статистика, когнитивная психология и дизайн. В условиях постоянного увеличения объёма и размерности визуализируемых данных [3], весьма актуальна проблема формирования концептуально новых моделей визуализации. Для разработки новых методик визуализации, необходимо знать общие особенности процесса их построения, а также уметь ориентироваться в уже существующих методиках, в том числе тех, которые используются вне сферы информационной безопасности, в которой авторы проводят интенсивные исследования и разрабатывают подсистему визуализации.

Зачастую новые модели систем визуализации приходят из областей искусства, медиа-дизайна, маркетинга, биоинформатики и т.п. Несмотря на отличие области применения, сами концепции отображения данных и представления сложных отношений между ними остаются прежними [4–5] (рисунок 1): в первую очередь осуществляется сбор данных, их анализ, фильтрация и только потом преобразование данных в графические примитивы из которых складывается графическая мо-

дель, которая, вместе с инструментами управления и отображения, представляет модель визуализации.

Рис. 1. Этапы построения визуализации

Работы, рассматривающие методики визуализации данных, как правило, уделяют недостаточно внимания целостности данного процесса, а именно, не рассматривают требуемые для визуализации аспекты, не упоминают при помощи каких инструментов и библиотек реализуется графическая модель или какие концептуальные инструменты могут расширить возможности той или иной графической модели, а также не рассматривают новые и уникальные графические модели.

Например, в [4] большое количество внимания уделяется этапу подготовки данных, а сами модели визуализации рассматриваются на уровне уже готовых инструментов. Этап подготовки данных безусловно является важным, однако, именно выбор модели визуализации является ключевым моментом для эффективного взаимодействия данных и пользователя. В [3] представлен обзор графических моделей, однако не упоминаются основные аспекты, исходя из которых, необходимо совершать выбор модели или разрабатывать свою собственную. В [7] также представлен обзор классических методик визуализации, но отсутствуют новые и уникальные концепции, построение которых на текущий момент актуально. Отдельно можно выделить обзоры из областей, не касающихся информационной безопасности. Например, графические модели в области социологии [8] могут быть успешно перенесены в область информационной безопасности, но перед этим должны быть преобразованы в соответствии с особенностями данных, касающихся информационной безопасности.

При построении или выборе модели важно понимать, как различные этапы и элементы процесса визуализации влияют на модель комплексно. В этом обзоре рассматриваются некоторые методологические примитивы на примере поэтапного построения модели визуализации с уже подготовленными данными, с целью сформировать комплексное видение процесса создания модели и влияющих на неё аспектов. Рассматриваемые примитивы можно классифицировать как:

– примитивы процесса визуализации - аспекты, которые влияют на процесс построения модели визуализации и с использованием которых разрабатывается начальная метамодель;

– примитивы графических моделей - основные принципы построения модели визуализации; часто, именно выбор графической концепции определяет формат взаимодействия данных с пользователем, а также определяет ограничения и возможности расширения этого взаимодействия;

– дополнительные инструменты - компоненты инструментов, расширяющие возможности графических моделей.

В статье рассматриваются основные методологические примитивы на примере поэтапного построения модели визуализации с заранее подготовленными данными, с целью сформировать комплексное видение процесса создания модели и влияющих на неё аспектов. Приводится классификация примитивов и их связи между собой в соответствии с этапами построения модели. В конце обзора также приведены библиотеки визуализации на популярных языках программирования Java и JavaScript, которые можно использовать для реализации и дополнения построенных графических моделей.

2. Примитивы процесса визуализации. Перед разработкой собственной или выбором готовой модели визуализации, важно понимать аспекты в рамках которых эта модель должна функционировать. Эти аспекты можно использовать как в качестве элементов шаблона при построении, так и в качестве инструментов выбора для анализа уже существующих систем. Их можно классифицировать на два класса:

– аспекты информативности - элементы, в соответствии с которыми необходимо строить или выбирать модель визуализации; аспекты информативности задают требования к модели визуализации, влияя на информативность;

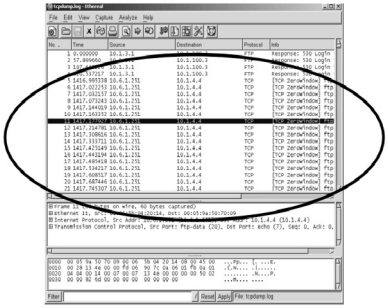
– аспекты эффективности - элементы, в соответствии с которыми желательно строить модель визуализации; аспекты эффективности задают требования к модели, влияя на эффективность представления информации пользователю.

Сами принципы и их наборы могут различаться в зависимости от цели визуализации и располагаемых ресурсов. Если предполагается что с механизмами визуализации будут работать пользователи, не обладающие необходимой квалификацией, то скорее всего, стоит уделить внимание аспектам, связанным с дизайном и, в целом, аспектам эффективного представления информации. В случае информационной безопасности, где точность и полнота информации имеют решающую роль, стоит уделить внимание аспектам информативности.

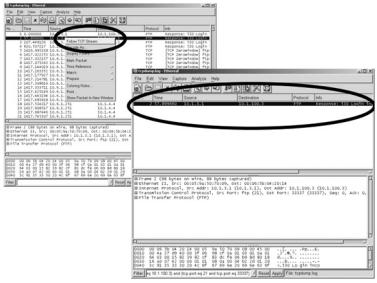
2.1. Примеры аспектов информативности. Аспекты информативности влияют на модель визуализации, задавая ограничения на количество информации, которую модель может отобразить. Рассмотрим несколько примеров аспектов, которые влияют на информативность модели.

2.1.1. Шаблон визуального поиска. Этот принцип был предложен Беном Шнейдерманом в 1996 году [9]. Согласно ему, визуальный поиск состоит из трёх этапов: (1) обзор ситуации в целом; (2) масштабирование и фильтрация; (3) детали по требованию. Соответственно, любая модель визуализации, цель которой не только представление, но и поиск, а также анализ информации, должна иметь инструменты для каждого из этапов.

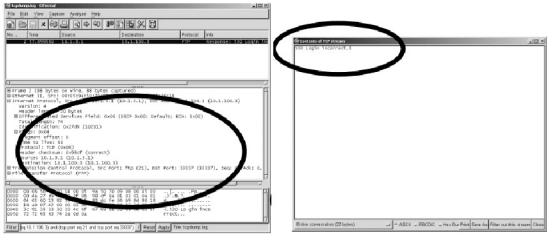
На рисунке 2 приведен пример выполнения данного принципа, на основе анализа перехваченных TCP-пакетов.



а) Обзор ситуации



б) Масштабирование и фильтрация



в) Детали по требованию

Рис. 2. Пример выполнения шаблона визуального поиска: а) обзор ситуации; б) масштабирование и фильтрация; в) детали по требованию

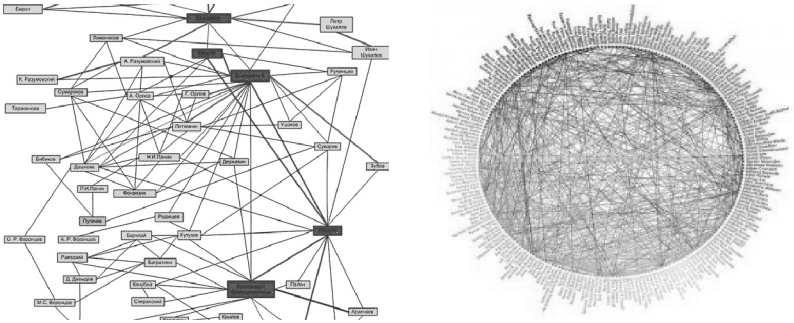
Изначально поиск осуществляется на верхнем уровне абстракции, позволяя видеть ситуацию по всем перехваченным пакетам в целом (рисунок 2а); на втором этапе пакеты фильтруются таким образом, что бы пользователь мог видеть пакеты в рамках конкретного соеди-

нения (рисунок 2б); на третьем этапе пользователь, по требованию, может перейти на нижний уровень абстракции - посмотреть подробности пакета (рисунок 2в).

2.1.2. Когнитивный аппарат. Эффективность модели визуализации целиком обусловлена когнитивным аппаратом пользователя (память, восприятие, ассоциации и т.д.). Модель визуализации должна учитывать эти особенности в целом и не выходить за пределы их возможностей [10]. Несмотря на то, что данный аспект целиком принадлежит области психологии, его базовые основы часто понятны на интуитивном уровне. Но чем больше и сложнее графическая модель, тем больше влияние данного аспекта. Особенно важно учитывать особенности когнитивного аппарата при разработке моделей, которые работают с большими или неоднородными данными. Примеры:

– с учётом психологических ассоциаций, красный (на рисунок 3а тёмный) цвет является “зарезервированным цветом” для отображения угроз;

– анализ отдельных элементов и метрик может быть затруднён, так как модель выходит за пределы возможностей когнитивного аппарата (пределы восприятия множества элементов) (рисунок 3б).



а) б)

Рис. 3. Графы, учитывающие особенности когнитивного аппарата: а) граф с выделением элементов “зарезервированными цветами”; б) сильно перегруженный элементами радиальный граф

2.1.3. Соответствие данных и их представления. При анализе данных необходимо, чтобы их визуализация соответствовала источнику. В некоторых моделях (например, использующих трёхмерное пространство) пропорциональное отношение между визуализацией и данными может быть нарушено.

В [11] Эдвард Тафти предлагает оценивать степень соответствия данных и изображенной информации при помощи коэффициента

“фактор лжи”, который равен $\frac{\text{эффективность визуализации}}{\text{эффективность данных}}$, где эффективность $= \frac{\text{значение a} - \text{значение b}}{\text{значение a}}$. Для эффективности визуализации выбираются графические представления двух значений. Для эффективности данных – данные, по которым строятся графические представления.

Рассмотрим пример: в инфографике (рисунок 4), цель которой показать на сколько увеличился расход топлива на человека в США, значение для 1978 года равно 18 галлонам, а для 1985 года - 27.5 галлонам, т.е. эффективность данных равна 53%; ширина дороги, которая является визуализацией расхода топлива, для 1978 года равна 1 сантиметру, для 1985 равна 7.8 сантиметрам; эффективность визуализации равна 780%; подобное несовпадение эффективности визуализации и эффективности данных даёт фактор лжи = 17.8 [11].

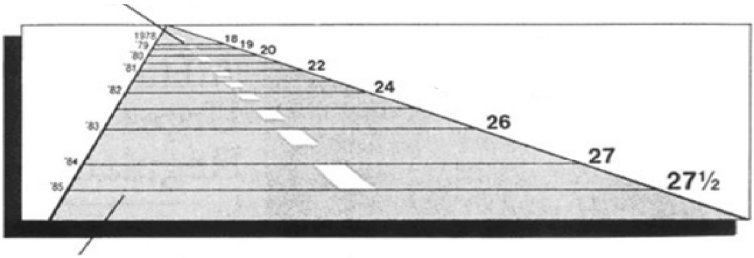


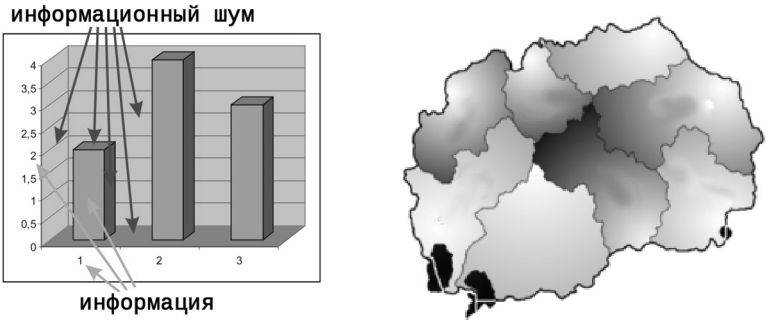
Рис. 4. Инфографика расхода топлива на человека в США

2.2. Примеры аспектов эффективности. Аспекты эффективности больше влияют на графическую модель, чем на модель визуализации. Они определяют эффективность диалога между механизмами визуализации и пользователем в рамках модели визуализации. Рассмотрим несколько аспектов, влияющих на эффективность.

2.2.1. Контроль информационного шума. Оптимальная модель не только та, в которую нечего добавить, но и из которой нечего убрать. Если элемент не входит в минимальный набор элементов визуализации, необходимых для решения поставленной задачи, он может являться источником информационного шума [12]. Как правило, подобные элементы являются элементами дизайна, который не был проработан в соответствии с особенностями модели визуализации.

На рисунке 5а модель становится более сложной для восприятия, за счёт необходимости анализировать данные в трёх измерениях. В данном случае информационным шумом является измерение, отображающее глубину.

На рисунке 5б пользователь может принять градиент внутри провинций в качестве параметра модели (например, плотность населения/температура и т.п.).



а) б)

Рис. 5. Модели визуализации с информационным шумом: а) трёхмерный график с информационным шумом; б) карта с информационным шумом в виде градиентов

2.2.2. Прямые манипуляции. Инструменты для взаимодействия между информацией и пользователем должны иметь аналоги в реальности [13]. Подобный подход использует когнитивный аппарат человека и позволяет пользователю предугадать, что произойдёт после взаимодействия, а также выбирать инструменты и работать с ними на уровне интуиции.

Так, например, в концепции тактильных поверхностей Material Design [14–15] каждый контейнер аналогичен листу бумаги; выбранный контейнер за счёт тени располагается к наблюдателю ближе по сравнению с остальными, давая понять на интуитивном уровне о его активности (рисунок 6).

Рис. 6. Тактильные поверхности Material Design

2.2.3. Графический дизайн. Модель визуализации должна работать в рамках одной из моделей графического дизайна или же иметь

свою собственную. Модель графического дизайна также можно представить в качестве элементов, каждый из которых должен иметь обоснование своего присутствия в модели. Часто именно графический дизайн имеет наибольшее влияние на эффективность взаимодействия модели визуализации и пользователя. Пример графика в рамках модели графического дизайна Material Design [14–15] изображён на рисунке 7.

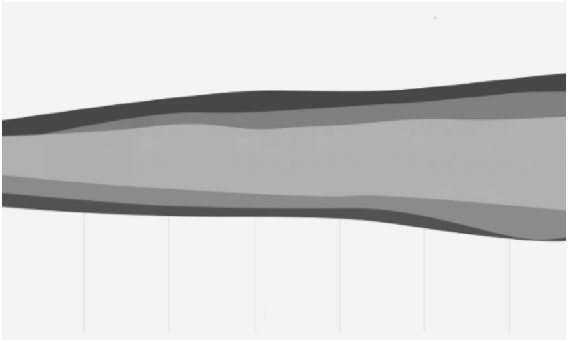


Рис. 7. График в рамках модели графического дизайна

3. Примитивы графических моделей. Модель визуализации является ядром процесса визуализации, определяя большую часть её ограничений и возможностей. Как правило, выбор модели напрямую зависит от цели процесса визуализации и типа информации. По степени использования, модели можно классифицировать как:

– широко распространенные;

– средне распространённые;

– специфичные.

3.1. Примеры широко распространённых моделей. К широко распространенным моделям, как правило, относят: графики, графы и их вариации и карты. Универсальность этих моделей обусловлена возможностью визуализации практически любых типов данных.

3.1.1. Графики. Графики получили широкое распространение как наиболее простая модель визуализации небольших данных [16]. При этом, степень успеха визуализации часто зависит от выбора графика, который, в свою очередь, зависит от типа данных [17-18]. На рисунке 8 изображены некоторые примеры простых графиков.

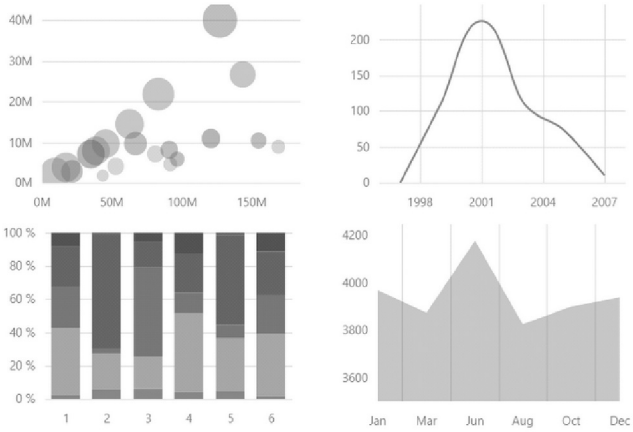


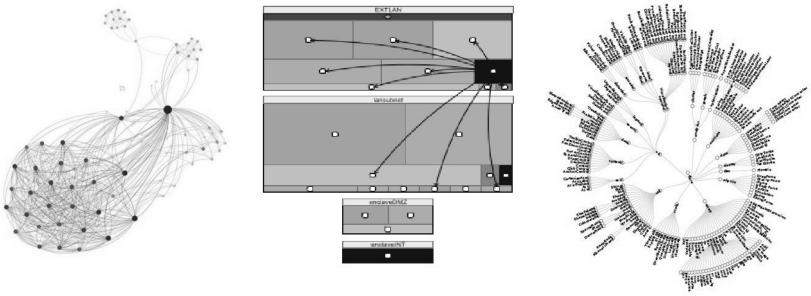
Рис. 8. Примеры простых графиков

3.1.2. Графы. Графы особенно популярны в информационной сфере: они интуитивно понятны, имеют множество вариаций и способны отображать большие объёмы разнородных данных [16]. На рисунок 9 изображены некоторые виды графов:

– классические - стандартные графы обычного представления (рисунок 9а);

– карты деревьев - иерархические графы на плоскости, вершины которых представлены прямоугольниками а отношения вложенностью (рисунок 9б);

– радиальные - иерархические графы, элементы которых расположены радиально (рисунок 9в).



а) б) в)

Рис. 9. Примеры различных графов: а) стандартный граф; б) карта дерева; в) радиальный граф

3.1.3. Карты. Карты используются в случаях, когда географические данные фигурируют в качестве ключевых [16]. При этом, остальные параметры выражаются в виде других графических моделей (графики, графы и т.д.), накладываясь на модель карт. На рисунке 10 представлена географическая карта атак [19–20] в реальном времени.

Рис. 10. Географическая карта атак

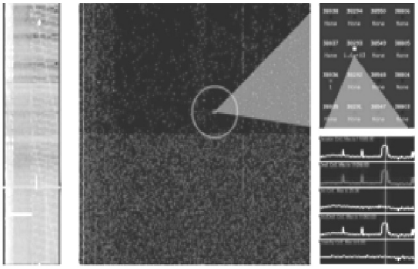
3.2. Примеры средне распространённых моделей. Средне распространенные модели - это модели, созданные в рамках определенной задачи, но имеющие потенциал в визуализации данных, выходящих за рамки цели их создания.

3.2.1. Матрицы. Для того что бы замаскировать атаку, атакующий обычно изменяет идентифицирующие его параметры, такие, например, как IP-адрес. Таким образом, чтобы идентифицировать нарушителя, необходимо опираться на другие параметры, например время прибытия пакета, которое зависит от типа операционной системы, задержки маршрутизатора и прочие метрики, которые трудно изменить.

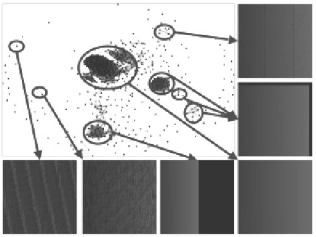
В [21] предлагается последовательный анализ на основе матричного представления.

На рисунке 11а в матрице слева, где время представлено вертикальной шкалой, один из всплесков фигурирует на протяжении нескольких часов. Матрица в центре показывает активность всех портов в выбранном временном диапазоне. Позиции всплеска на матрице слева соответствует всего один аномально активный порт в центральной матрице. Графики справа показывают, что большому количеству получателей соответствует небольшое количество источников, что, вероятнее всего, является признаком сканирования сети.

В [21] также предлагается использование данной модели в совокупности с классическими графами для выборочного анализа кластеров сети (рисунок 11б).



а)

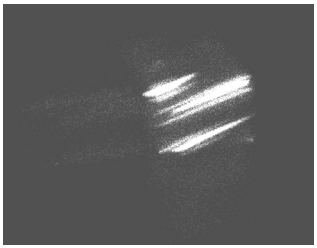


б)

Рис. 11. Примеры моделей на основе матриц: а) модель на основе матричного представления; б) использование матриц для анализа кластеров

В [22] анализируется распределение TCP/IP-адресов с целью уменьшения рисков TCP/IP-спуфинга. Несмотря на то, что операционные системы используют для распределения генераторы случайных чисел, при их визуализации, несмотря на высокую или полную случайность, для каждой операционной системы можно выделить кластеры IP-адресов, наличие которых может помочь нарушителю.

На рисунке 12а изображен анализ последовательности адресов созданных операционной системой UNICOS 10.0.0.8. Несмотря на низкую оценку целесообразности атаки со стороны операционной системы (рисунок 12б), можно выделить 3 больших кластера, в пределах которых атака наверняка будет более успешной.



а)

OS Name:	UNICOS 10.0.0.8
R1 radius:	10
Average R2:	948
Average N:	47
Average error:	390
Attack feasibility:	9.00%

б)

Рис. 12. Пример использования матриц для анализа распределения: а) матрица распределения TCP/IP-адресов; б) вероятностная оценка целесообразности атаки

3.2.2. Гистограммы. При необходимости сделать общие выводы, матричное представление можно преобразовать в гистограммы. Матрица на рисунке 13 получена с использованием параметра, рассчитанного на основе анализа посещений с уникальным адресом, где красный (на рисунке 13 - светлый) цвет соответствует максимуму количества, а черный - отсутствию посещений [21].

В гистограммах верхних матриц сравниваются похожие всплески на предмет различий. В гистограммах нижних матриц сравниваются различные всплески на предмет совпадений.

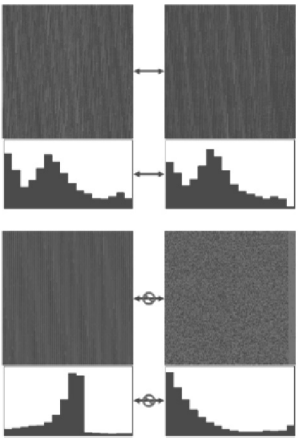
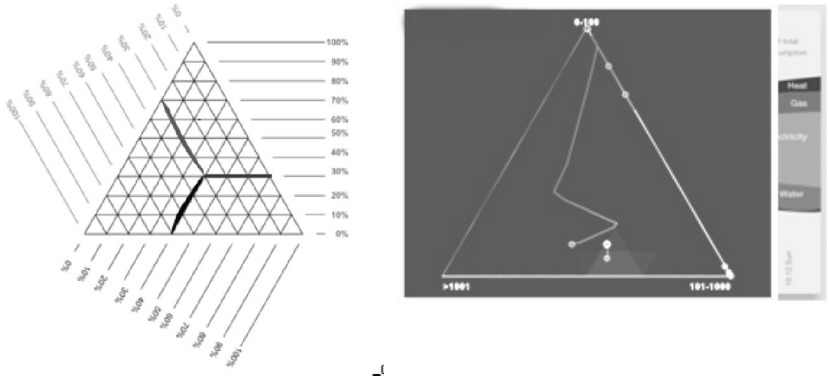


Рис. 13. Представление матриц в виде гистограмм

3.2.3. Трилинейные координаты. В [1] предлагается модель визуализации (рисунке 14б), разработанная на основе треугольной модели визуализации, которая была представлена геологическим департаментом США. Модель оперирует тремя метриками, каждая из которых соотносится с ребром треугольника и выражается в процентом соотношении по отношению друг к другу. В качестве примера, было взято процентное соотношение сообщений, переданных по различным протоколам (TCP, UDP, ICMP). Возможные положения источника в трилинейных координатах, а также положение источника с параметрами 30%, 40% и 30% приведены на рисунке 14а.



а) б)

Рис. 14. Пример модели на основе трилинейных координат: а) возможные положения источника в трилинейных координатах; б) Модель на основе трилинейных координат

Для отображения изменений во времени, за каждым перемещением источника прокладывается трасса (рисунок 14б). Зоны, в которых источник проводит больше всего времени, подкрашиваются синим (на рисунок 14б - светлым) цветом.

Модель позволяет выявить аномальное поведение источника, когда он переходит в область, не типичную для его местонахождения, либо же когда движение для него само по себе аномально.

3.2.4. Параллельные координаты. Модель параллельных координат [23] является частным случаем графика. Она позволяет эффективно отобразить многомерные данные, располагая каждый тип данных вдоль одной из параллелей (рисунок 15).

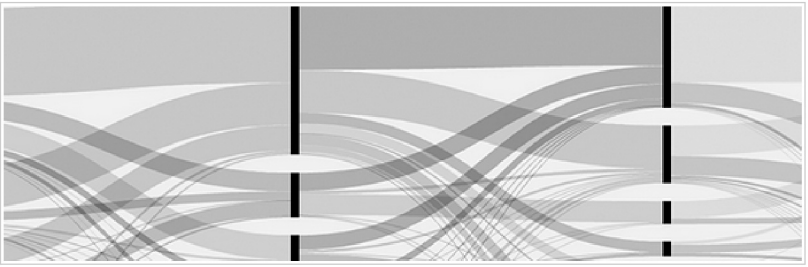


Рис. 15. Модель на основе параллельных координат

В [24] предлагается похожая модель, оси которой расположены не параллельно, а радиально (рисунок 16). Предполагается, что использование полярных координат позволит лучше выявлять аномалии на когнитивном уровне, а также отображать метрики с большей эффективностью.

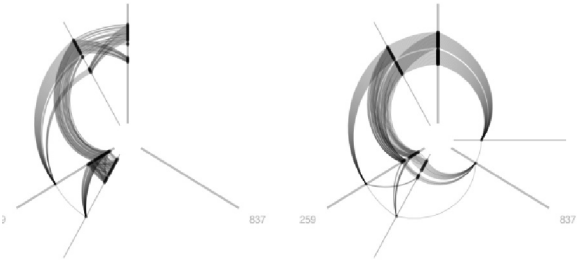


Рис. 16. Модель радиальных координат

3.3. Примеры специфичных моделей. Специфичные модели - это модели специализированного типа, созданные для решения конкретных задач, и которые, как правило, не могут быть использованы для решения задач другого вида.

3.3.1. Визуализация инструментов работы с лог-файлами. В [25] предлагается визуализация инструментов для детальной работы с лог-файлами на основе регулярных выражений (рисунок 17). Преимуществами данного подхода являются: возможность работы с абстракцией (визуализацией) и самими лог-файлами одновременно, быстрая навигация без потери ориентации, анализ подмножеств записей с учётом контекста и исследовательский характер поиска, вместо стандартных поисковых запросов.

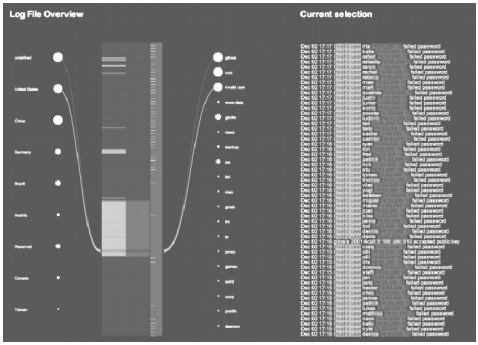
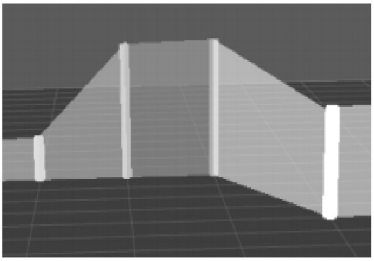


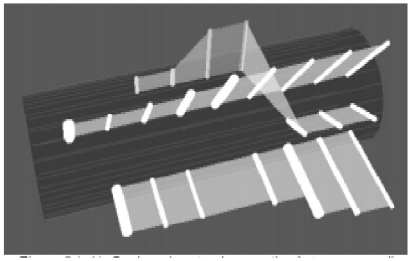
Рис. 17. Визуализация инструментов для работы с лог файлами

3.3.2. Визуализация сложных атак. В [26] предлагается трехмерная модель визуализации сложных многоступенчатых атак (рисунок 18).

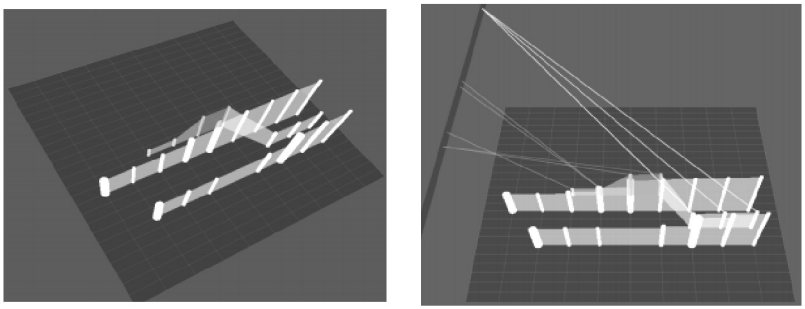
Каждый цилиндр представляет собой событие, тип которого задаётся цветом: зелёный - сканирование, фиолетовый - попытка удалённого доступа, красный - успешная попытка удалённого доступа, жёлтый - DoS (на рисунке 18 цвета заданны соответствующими оттеками серого).



а)



б)



в) г)

Рис. 18. Модель визуализации сложных атак: а) параметры разных видов атак; б) модель в полярных координатах; в) последовательное расположение атак; г) отображение источников атак в виде отдельной плоскости

Уровень серьёзности оповещений задаётся высотой цилиндра (рисунок 18а).

Ключевая особенность системы - это возможность постепенного отслеживания атаки как единого целого, зная какие события предшествовали ей и какие произошли после (рисунок 18в).

При этом система обладает рядом дополнительных инструментов, таких как переход от декартовых к полярным координатам (рисунок 18б), возможность отображения источников атак в соответствии с типом атаки (рисунок 18г), а также управление плоскостью источников для обеспечения лучшего угла обзора (рисунок 18г).

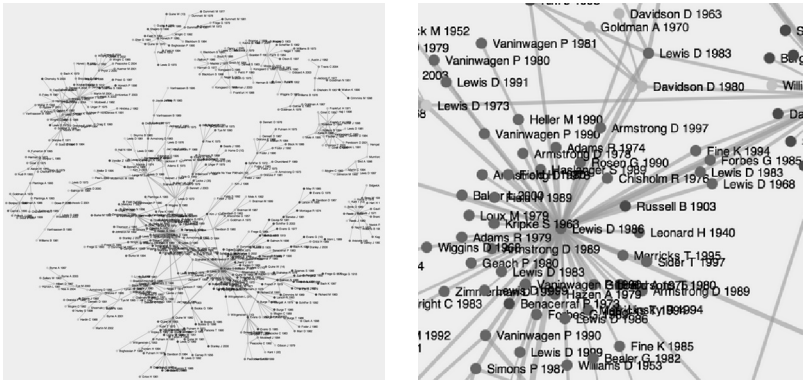
4. Дополнительные инструменты. Для некоторых типов данных или определенного набора данных, выбор графической модели может быть представлен всего несколькими видами моделей или вовсе отсутствовать.

Например, для представления топологии сети наглядным вариантом всегда является представление в виде графа.

При этом, если многомерность данных или их параметров сопровождается большим объёмом, модель визуализации в определенные моменты может выходить за ограничительные пределы процесса визуализации.

Наиболее остро данная проблема стоит при визуализации больших сетей, когда пользователь вынужден ещё на стадии обзора жертвовать отображением одних метрик, для отображения других.

На рисунке 19, для отображения отдельных элементов (рисунок 19б), пользователь, воспользовавшись инструментом масштабирования, вынужден отказаться от отображения сети целиком (рисунок 19а).

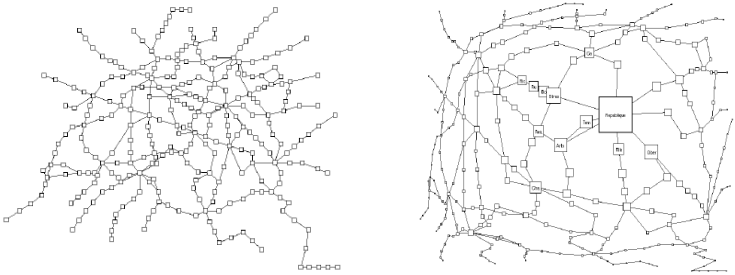


а) б)

Рис. 19. Пример визуализации большой сети: а) перегруженный граф; б) увеличенный сегмент графа

Таким образом, можно выделить отдельный класс примитивов - инструменты, которые не могут являться самостоятельными графическими моделями, но могут расширить уже существующую модель, не нарушая её принципов.

4.1. Инструмент “Рыбий глаз”. Основным параметром, влияющим на эффективность восприятия при визуализации графов, является количество узлов и связей между ними. Классические пути решения проблемы с количеством узлов и связей - это переход на более высокий уровень абстракции (что требует разработки модели представления более высокого уровня, либо же связей между уровнями), либо масштабирование, в процессе которого увеличивается одна часть графа и полностью теряется другая, нарушая видение ситуации в целом. В [27] Саркар и Браун предлагают использовать вместо масштабирования эффект “рыбьего глаза” (рисунок 20).



а) б)

Рис. 20. Пример использования инструмента "Рыбий глаз": а) граф без применения линзы; б) граф с применением линзы

Предполагается, что пользователь способен выбрать область фокуса, в результате чего интересующие его элементы станут больше, в то время как остальные также останутся в поле зрения. В отличии от обычного масштабирования, пользователю не надо мысленно объединять изображения для полного обзора. В инструменте "рыбий глаз" кратности масштаба соответствует коэффициент искажения.

4.2. Инструмент "Множественный взгляд". Для некоторых типов данных полезно использование сразу нескольких моделей визуализации одновременно [28]. Нахождение необходимых отношений или аномалий при данном подходе значительно упрощено. При многомерности метрик, подход, отображающий для каждого множества параметров соответствующую оптимальную графическую модель, может оказаться эффективнее стандартных средств, таких как переход между уровнями абстракции или исключение ряда параметров.

На рисунке 21 один и тот же набор данных представлен параллельными координатами, 2D графом и картой деревьев.

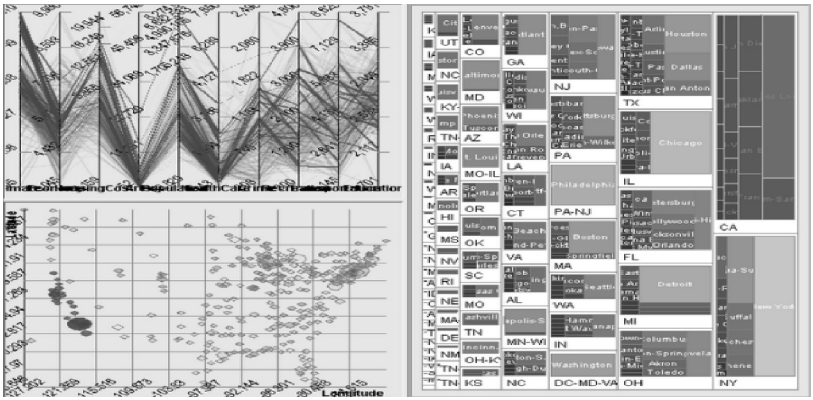


Рис. 21. Различные представления одного графа

4.3. Инструмент "Семантическое масштабирование". Восприятие графических форм в некоторых случаях менее эффективно семантического восприятия. Таким образом, при работе с представлением в виде текста также необходимо иметь соответствующие инструменты. Масштабирование возможно не только на уровне абстракций или отдельно взятого изображения, но и на уровне семантики [29]. Классическими примером семантического масштабирования является текстовая каталогизация, когда элементы группируются не по функциональным, а по семантическим параметрам, таким как алфавитный порядок, лексическое значение и т.д.

4.4. Инструмент “Небольшие различия”. Иногда, при применении фильтров, необходимо видеть не только итог фильтрации, но и её предварительный результат, или несколько результатов сразу. Например, на рисунке 22 в левой части отображено множество результатов применений фильтра с разными параметрами. Маленькие, размером с почтовую марку компоненты, предназначенные для моментального ответа на вопрос “а что если?”, могут не только ускорить процесс поиска информации, но и избавить от необходимости однозначно выбирать модель визуализации, где в качестве одной из метрик будет предварительный результат [30–31].

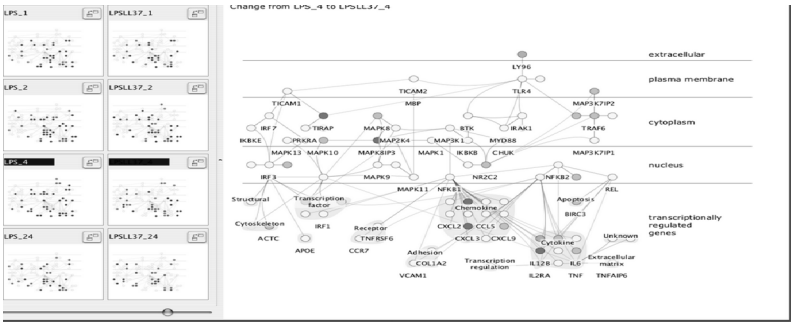


Рис. 22. Отображение вариаций применения фильтра

4. Библиотеки визуализации. Построение модели визуализации всегда зависит от метрик, некоторые из которых часто встречаются во многих уже существующих системах. Таким образом, может получится, что разработчик будет реализовать то, что было разработано другими специалистами ранее.

Знание основных библиотек для визуализации данных и шаблонов реализации на их основе однозначно необходимо перед построением модели визуализации. Также библиотеки содержат уже готовые реализации некоторых компонентов системы и предоставляют различные фреймворки для упрощения процесса разработки.

Ниже приведено несколько популярных некоммерческих библиотек для языков Java и JavaScript.

4.1. Jung. Простая и стабильная библиотека Jung [32] разработана на Java и используется преимущественно для визуализации графов.

Среди недостатков этой библиотеки можно отметить, что разработчиками не была должным образом проработана документация, а также то, что проект приостановил своё развитие.

Таким образом, Jung может быть полезна только для визуализации простых графов (рисунок 23).

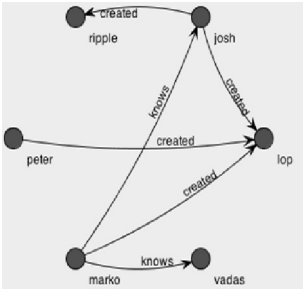


Рис. 23. Пример графа, реализованного на Jung

4.2. GraphViz. Визуализация в рамках пакета GraphViz [33] ведётся при помощи языка описания графов DOT [33]. GraphViz принимает файл на языке DOT и автоматически формирует изображение с заранее определенной моделью.

Из достоинств данной библиотеки можно отметить хорошо проработанную документацию, простоту использования, наличие возможности кластеризации и поддержку нескольких графических моделей. Из недостатков можно отметить, что GraphViz может использоваться только для визуализации графов.

Пример визуализации графа с разбиением на кластеры приведен на рисунке 24.

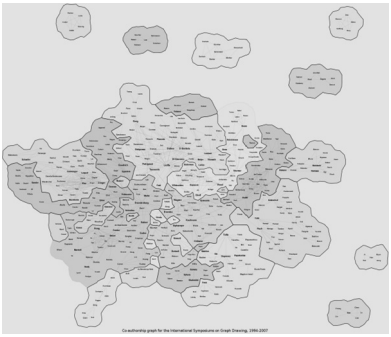


Рис. 24. Пример графа реализованного на GraphViz

4.3. Prefuse. Данная библиотека использует стандартную Java библиотеку Java2D, поэтому она легко интегрируется с приложениями, разработанными с использованием пакета Java Swing. Prefuse [34] содержит широкий выбор инструментов (в том числе подключение к базе

данных, наличие собственных структур данных, поддержка анимации, динамических запросов, поиска, таблиц, графиков и т.д.) и, как следствие, предназначена не только для визуализации графов. Prefuse является достаточно гибким средством, однако ее развитие приостановлено, вследствие чего документация проработана недостаточно хорошо. Пример графа с использованием Prefuse изображен на рисунке 25.

Рис. 25. Пример графа реализованного на Prefuse

4.4. D3. Эта библиотека [35] разработана на JavaScript и является одной из наиболее популярных библиотек для визуализации данных. Её можно использовать как основу для реализации модели визуализации в рамках фреймворка Data-Driven-Document, или же использовать одну из сотен готовых реализаций графических моделей, предлагаемых разработчиками. Данная библиотека имеет хорошо проработанную документацию, причем эта документация переводится на множество языков, в том числе и на русский. Примеры графических моделей, реализованных на D3, изображены на рисунке 26.

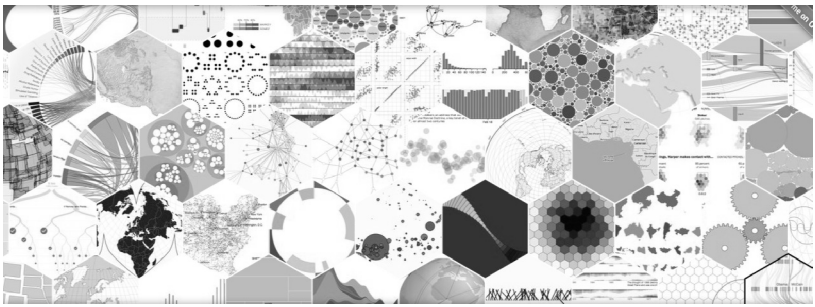


Рис. 26. Примеры различных моделей, реализованных на D3

5. Заключение. Можно сказать, что целью процесса визуализации является нахождение компромисса между показателями информативности данных и эффективности их представления. Каждый элемент модели визуализации - аспекты процесса визуализации, графическая модель, инструмент, а также библиотека реализации - по отдельности влияют на эти показатели. Но в конечном итоге, решение об использовании тех или иных элементов зависит именно от данных. Здесь важно понимать что не данные должны определяться моделью визуализации, а модель визуализации должна формироваться на основе данных. Порой, более правильным набором элементов будет простой и минимальный набор, чем сложная графическая модель с большим количеством инструментов под управлением множества примитивов.

В данном обзоре приведены основные элементы процесса визуализации. Знание этих элементов для анализа уже существующей модели и умение их применять на разных этапах разработки модели визуализации поможет, если не достигнуть компромисса информативности и эффективности, то максимально к нему приблизиться.

Описанные в данной статье методы успешно применяются авторами как для визуализации общих моделей атак и метрик защищенности [36], так и для визуализации отдельных элементов атак представленных в базе данных CAPEC [20].

Литература

1. *Bruce R.* Applying Information Visualization to Computer Security Applications // All Graduate Theses and Dissertations. 2010. 636 p.
2. *Ботя М.В.* Инфографика как объект информационного дизайна // Новые информационные технологии в образовании: материалы VIII международной научно-практической конференции. Екатеринбург, 2015. С. 411–414.
3. *Клышинский Э.С., Рысаков С.В. и Шихов А.И.* Обзор методов визуализации многомерных данных // Новые информационное технологии в автоматизированных системах. 2014. С. 519–530.
4. *Falschlunger L., Lehner O., Eisl C., Losbichler H.* Development of a Data Visualization Model based on Information Processing Theory // Proceedings of the 9th conference for Austrian universities of applied sciences. Hagenberg. Austria. 2015. pp. 1–7.
5. *Novikova E., Kotenko I.* Analytical Visualization Techniques for Security Information and Event Management // 21st Euromicro International Conference on Parallel, Distributed, and Network-Based Processing. 2013. pp. 519–525.
6. *Ferebee D. Dasgupta D.* Security Visualization Survey // Proceedings of the 12th Colloquium for Information Systems Security Education University of Texas. 2008. 124 p.
7. *Barrera D.* Towards Classifying And Selecting Appropriate Security Visualisation Techniques // A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master Of Science School of Computer Science at Carleton University Ottawa. Ontario. 2009. 117 p.
8. *Healy K., Moody J.* Data Visualization in Sociology // Annual Review of Ecology and Systematics 2014. 2014. pp. 105–128.

9. *Shneiderman B.* The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations // In Proceedings of the IEEE Symposium on Visual Languages. Washington. 2006. pp. 336–343.
10. *Goldstein B.* Cognitive Psychology // Thomson Wadsworth. 2005.
11. *Tufte* E. The Visual Display of Quantitative Information // Graphics Press. USA. 1991. pp. 57–69.
12. *Tufte E.* Visual Explanations // Graphics Press. Cheshire. Connecticut. 1997.
13. *Hutchins E., Hollan J., Norman D.* Direct Manipulation Interfaces // Lawrence Erlbaum Associates. 1985. vol. 1. pp. 311–338.
14. *Google inc.* Material-Design Introduction. URL: www.google.com/design (дата обращения 01.10.15).
15. Блог компании REDMADROBOT. URL: www.habrahabr.ru/company/redmadrobot (дата обращения 01.10.15).
16. *Marty R.* Applied Security Visualization // Addison Wesley Professional. 2009.
17. *Kotenko I., Novikova E.* Visualization of Security Metrics for Cyber Situation Awareness // The 1st International Software Assurance Workshop (SAW 2014). In conjunction with the 9th International Conference on Availability, Reliability and Security (ARES 2014). Fribourg, Switzerland. IEEE Computer Society. 2014. pp. 506–513.
18. Блог компании Devexpress. URL: www.habrahabr.ru/company/devexpress (дата обращения 01.10.15).
19. Сайт компании Kaspersky с презентацией карты атак реального времени. URL: www.cybermap.kaspersky.com (дата обращения 01.10.15).
20. *Котенко И.В., Дойникова Е.В., Чечулин А.А.* Общее перечисление и классификация шаблонов атак (CAPEC): описание и примеры применения // Защита информации. Инсайд. 2012. №4. С. 54–66.
21. *Kwan-Liu M.* Cyber Security Through Visualization // In Proc. Asia Pacific Symposium on Information Visualisation. Tokyo. Japan. 2006.
22. *Zalewski M.* TCP/IP Sequence Number Analysis. 2002. URL: www.lcamtuf.coredump.cx/newtcp. (дата обращения 01.10.15).
23. *Tricaud S.* Visualizing Network Activity using Parallel Coordinates // System Sciences. 44th Hawaii International Conference. 2011. pp. 1–8.
24. *Engle S., Whalen S.* Visualizing Distributed Memory Computations with Hive Plots // VizSec '12 Proceedings of the Ninth International Symposium on Visualization for Cyber Security. 2012.
25. *Stange J., Dörk J., Landstorfer J., Wettach R.* Visual Filter: Graphical Exploration of Network Security Log Files // VizSec '14 Proceedings of the Eleventh Workshop on Visualization for Cyber Security. 2014. pp. 41–48.
26. *Yelizarov A., Gamayunov D.* Visualization of Complex Attacks and State of Attacked Network // Visualization for Cyber Security. VizSec 2009: 6th International Workshop. 2009. pp. 1–9.
27. *Sarkar M., Brown M.* Graphical fisheye views // Communications of the ACM. 1994. vol. 37. no. 12. pp. 73–83.
28. *Wang M., Woodruff A., Kuchinsky A.* Guidelines for Using Multiple Views in Information Visualization // Proc. of Advanced Visual Interfaces. 2000. pp. 110–119.
29. *Watson G.* Lecture Lecture 15 - Visualisation of Abstract Information // Edinburgh Virtual Environment Centre. 2004.
30. *Wroblewski L.* Small Multiples Within a User Interface // Web Form Design. 2005.
31. *Tufte E.* Envisioning Information // Graphics Press. Cheshire. 1990.
32. Официальный сайт библиотеки Jung. URL: www.jung.sourceforge.net (дата обращения 01.10.15).
33. Официальный сайт библиотеки Graphviz. URL: www.graphviz.org (дата обращения 01.10.15).

34. Официальный сайт библиотеки Prefuse. URL: www.prefuse.org (дата обращения 01.10.15).
35. Официальный сайт библиотеки D3js. URL: www.d3js.org (дата обращения 01.10.15).
36. *Kotenko I.V., Chechulin A.A.* A Cyber Attack Modeling and Impact Assessment Framework // Proceedings of the 5th International Conference on Cyber Conflict 2013 (CyCon 2013). IEEE and NATO COE Publications. Tallinn. Estonia. 2013. pp. 119–142.

References

1. Bruce R. Applying Information Visualization to Computer Security Applications // All Graduate Theses and Dissertations. 2010. 636 p.
2. Botja M.V. [Infographic as Object of Information Design]. *Novye informacionnye tehnologii v obrazovanii: materialy VIII mezhdunarodnoj nauchno-prakticheskoj konferencii* [New information technologies in education: materials of VIII scientific-practice conference]. Ekaterinburg. 2015. pp. 411–414. (In Russ.).
3. 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.).
4. Falschlunger L., Lehner O., Eisl C., Losbichler H. Development of a Data Visualization Model based on Information Processing Theory. Proceedings of the 9th conference for Austrian universities of applied sciences. Hagenberg. Austria. 2015. pp. 1–7.
5. Novikova E., Kotenko I. Analytical Visualization Techniques for Security Information and Event Management. 21st Euromicro International Conference on Parallel, Distributed, and Network-Based Processing. 2013. pp. 519–525.
6. Ferebee D. and Dasgupta D. Security Visualization Survey. Proceedings of the 12th Colloquium for Information Systems Security Education University of Texas. 2008. 124 p.
7. Barrera D. Towards Classifying And Selecting Appropriate Security Visualisation Techniques. A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master Of Science School of Computer Science at Carleton University Ottawa. Ontario. 2009. 117 p.
8. Healy K., Moody J. Data Visualization in Sociology. *Annual Review of Ecology and Systematics*. 2014. pp. 105–128.
9. Shneiderman B. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In Proceedings of the IEEE Symposium on Visual Languages. Washington. 2006. pp. 336–343.
10. Goldstein B. Cognitive Psychology. Thomson Wadsworth. 2005.
11. Tufte E. The Visual Display of Quantitative Information. *Graphics Press*. USA. 1991. pp. 57 – 69.
12. Edward R. Tufte, Visual Explanations. *Graphics Press*. Cheshire. Connecticut. 1997.
13. Hutchins E., Hollan J., Norman D. Direct Manipulation Interfaces. Lawrence Erlbaum Associates. 1985. vol. 1. pp. 311–338.
14. Google inc. Material-Design Introduction. Available at: www.google.com/design. (accessed 01.10.2015).
15. Blog kompanii REDMADROBOT [REDMADROBOT company blog]. Available at: www.habrahabr.ru/company/redmadrobot. (accessed 01.10.2015). (In Russ.).
16. Marty R. Applied Security Visualization. Addison Wesley Professional. 2009.
17. Kotenko I., Novikova E. Visualization of Security Metrics for Cyber Situation Awareness. The 1st International Software Assurance Workshop (SAW 2014). In con-

junction with the 9th International Conference on Availability, Reliability and Security (ARES 2014). Fribourg. Switzerland. 2014. pp. 506–513.
18. Blog kompanii Devexpress [Devexpress company blog]. Available at: www.habrahabr.ru/company/devexpress. (accessed 01.10.2015). (In Russ.).
19. Sajt kompanii Kaspersky s prezentaciej karty atak realnogo vremeni [Kaspersy web site with real time attack cybermap]. Available at: www.cybermap.kaspersky.com. (accessed 01.10.2015).
20. Kotenko I.V., Dojnikova E.V., Chechulin A.A. [General enumeration and classification of attack patterns (CAPEC): description and application examples]. *Zashhita informacii. Insajd – Data protection. Inside*. 2012. vol. 4. pp. 54–66.
21. Kwan-Liu M. Cyber Security Through Visualization, In Proc. Asia Pacific Symposium on Information Visualisation. Tokyo. Japan. 2006.
22. Zalewski M. TCP/IP Sequence Number Analysis. 2002. Available at: www.lcamtuf.coredump.cx/newtcp. (accessed 01.10.2015).
23. Tricaud S. Visualizing Network Activity using Parallel Coordinates, System Sciences. 44th Hawaii International Conference. 2011. pp. 1–8.
24. Engle S., Whalen S. Visualizing Distributed Memory Computations with Hive Plots. VizSec '12 Proceedings of the Ninth International Symposium on Visualization for Cyber Security. 2012.
25. Stange J., Dörk J., Landstorfer J., Wettach R. Visual Filter: Graphical Exploration of Network Security Log Files. VizSec '14 Proceedings of the Eleventh Workshop on Visualization for Cyber Security. 2014. pp. 41–48.
26. Yelizarov A., Gamayunov D. Visualization of Complex Attacks and State of Attacked Network, Visualization for Cyber Security. VizSec 2009: 6th International Workshop. 2009. pp. 1–9.
27. Sarkar M., Brown M. Graphical fisheye views. Communications of the ACM. 1994. vol. 37. no. 12. pp. 73–83.
28. Wang M., Woodruff A., Kuchinsky A. Guidelines for Using Multiple Views in Information Visualization. Proc. of Advanced Visual Interfaces. 2000. pp. 110–119.
29. Watson G. Lecture Lecture 15 - Visualisation of Abstract Information. Edinburgh Virtual Environment Centre. 2004.
30. Wroblewski L. Small Multiples Within a User Interface, Web Form Design. 2005.
31. Tufte E. Envisioning Information. *Graphics Press*. Cheshire. 1990.
32. Oficial'nyj sajt biblioteki Jung [Official web site of Jung library]. Available at: www.jung.sourceforge.net. (accessed 01.10.2015).
33. Oficial'nyj sajt biblioteki Graphviz [Official web site of Graphviz library]. Available at: www.graphviz.org. (accessed 01.10.2015).
34. Oficial'nyj sajt biblioteki Prefuse library [Official web site of Prefuse library]. Available at: www.prefuse.org. (accessed 01.10.2015).
35. Oficial'nyj sajt biblioteki D3js [Official web site of D3js library]. Available at: www.d3js.org. (accessed 01.10.2015).
36. Kotenko I.V., Chechulin A.A. A Cyber Attack Modeling and Impact Assessment Framework. Proceedings of the 5th International Conference on Cyber Conflict 2013 (CyCon 2013). IEEE and NATO COE Publications. Tallinn. Estonia. 2013. pp. 119–142.

Коломеец Максим Вадимович — программист 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 — 0. guardeecwalker@gmail.com; 14-я ли-

ния В.О., д. 39, ком. 205, Санкт-Петербург, 199178; р.т.: +7(812)328–2642, Факс: +7(812)328–4450.

Kolomeec Maxim Vadimovich — developer 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 — 0. guardeecwalker@gmail.com; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7(812)328–2642, Fax: +7(812)328–4450.

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 — 110. 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 — 110. 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., д. 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 (13-01-00843, 14-07-00697, 14-07-00417, 15-07-07451, 15-37-51126) и за счет гранта РНФ 15-11-30029 в СПИИРАН

Acknowledgements. This research is supported by RFBR (projects no. 13-01-00843, 14-07-00697, 14-07-00417, 15-07-07451, 15-37-51126) as well as by RSF project 15-11-30029.

РЕФЕРАТ

Коломеец М.В., Чечулин А.А., Котенко И.В. **Обзор методологических примитивов для поэтапного построения модели визуализации данных.**

В условиях постоянного увеличения объёма и размерности информации, весьма актуальна проблема разработки концептуально новых моделей визуализации. Построение модели визуализации данных это сложный процесс, на который влияют большое количество аспектов. Для разработки новых методик, необходимо знать общие особенности процесса их построения, а так же уметь ориентироваться в уже существующих методиках. При этом, важно понимать, как различные этапы и элементы процесса визуализации влияют на модель комплексно.

В статье рассматриваются основные методологические примитивы на примере поэтапного построения модели визуализации с заранее подготовленными данными. Приводится классификация примитивов и их связей в соответствии с этапами построения модели. Рассматриваются библиотеки визуализации на популярных языках программирования.

SUMMARY

Kolomeec M.V., Chechulin A.A., Kotenko I.V. **Review of Methodological Primitives for the Phased Construction of Data Visualization Model.**

Development of new conceptual visualization models is an actual problem, especially when information volume and its dimensions are always growing up. Construction of data visualization model is a difficult process, which is influenced by different aspects. It is important to know common features of processes and already existing methods for constructing new visualization techniques. In addition, it is necessary to understand how different phases and elements influence visualization model comprehensively.

The paper considers common methodological primitives by the example of phased data visualization model construction with prepared data. The primitives classification and communication between them are suggested. The paper also considers data visualization libraries in different programming languages.