УДК 681.3.06(075.8)

Е. С. Федотов
Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В. И. Ульянова (Ленина)

Разработка методик визуального анализа для обнаружения внутреннего нарушителя

Рассматривается задача разработки методики визуального анализа для обнаружения внутреннего нарушителя. Объектом исследования является процесс выявления аномальной активности в поведении пользователя. Рассматривается методика построения визуального профиля пользователя. Приведен пример выявления аномальной активности и ее анализ.

Внутренний нарушитель, визуальный анализ, анализ логов системы, выявление аномалий

Проблема обнаружения внутреннего нарушителя относится к сложным задачам информационной безопасности. Под внутренним нарушителем понимается действующий или бывший сотрудник компании, который имеет или имел авторизированный доступ к сети организации или системе, с помощью доступа воздействует на конфиденциальность, целостность и доступность информационных ресурсов организации.

В настоящее время можно выделить несколько подходов к анализу контролируемых данных, предназначенных для обнаружения и предсказания угроз. В их основе лежат методы машинного обучения, поведенческого анализа и сигнатурного анализа [1], [2]. Несмотря на наличие достаточно большого количества программных средств, регистрирующих и контролирующих деятельность пользователей информационной системы, возможности по выявлению внутреннего нарушителя до сих пор достаточно ограничены, и с каждым годом их обнаружение становится сложнее, поскольку злонамеренные действия, на первый взгляд, неотличимы от легитимных, появляются новые сценарии атак, растет уровень навыков пользователей [3]. Возникают ситуации, когда необходимо анализировать разнообразные логи системных приложений для выявления нарушений безопасности и определения их причин. Изучение текстовых данных даже в табличном виде является трудоемкой задачей. Применение методов визуального анализа данных может значительно повысить эффективность работы аналитика за счет использования способностей человека быстро выявлять графические шаблоны и применения автоматизированных методов интеллектуальной обработки данных [4]. В настоящей статье описывается подход к визуальному анализу системных параметров, отличительной чертой которого является возможность сформировать профиль пользователя за заданный интервал времени и оценить его со статистическим профилем группы пользователей, имеющих ту же роль в информационной системе. Эффективность применения разработанной методики визуального анализа оценена на примере атаки злоумышленника на локальный компьютер пользователя во время обеда, когда сам пользователь отсутствовал на рабочем месте и не мог обнаружить данную атаку, а также предотвратить ее.

Современные специализированные системы обнаружения внутреннего нарушителя предлагают достаточно ограниченные возможности по визуальному анализу данных, используя, в основном, стандартные модели графического представления данных, такие, как секторные диаграммы, гистограммы и линейные графики[1]. С их помощью отображаются статистические данные о том,

[1] Spector 360 Web site, http://www.spector360.com/features/, 01.06.2015; StealthWatch, Lancope, https://www.lancope.com/threats/, 01.06.2015.

какие сервисы и программы наиболее часто используются в компании, какие веб-сайты чаще всего посещаются, какой сотрудник получает/отправляет больше всех писем и т. д. Системы управления информацией и событиями безопасности в большей степени ориентированы на обеспечение сетевой безопасности, анализируя сетевую активность пользователя, и реализуют схожий с системами обнаружения внутреннего нарушителя графический функционал[2]. В [5] предложено представлять активность пользователей в виде графа, вершинами которого являются пользователи системы и признаки аномальной активности. Недостатком данного подхода является необходимость заранее определить, что является вредоносной активностью, и в результате существует вероятность пропустить новые сценарии атак. В [6] описан инструмент для анализа текстовых данных, хранящихся на жестком диске, в котором для частотного анализа слов в тексте используются методики визуализации «облако слов» и карта деревьев для отображения числа ключевых слов в файле.

Для получения более точной оценки поведения пользователя с целью выявления аномальной активности необходимо исследовать данные, способные описать его деятельность наиболее полно: вход/выход в информационную систему, доступ к устройствам и ресурсам системы, использование внешних носителей, получение и отправку электронных писем. Это объясняется тем, что пользователь может проявлять подозрительную активность по одной из категорий данных, например доступ к файлам, а по другим – нет. Кроме того, анализ изменений в поведении пользователя следует выполнять в контексте среднестатистического поведения пользователей, имеющих одну и ту же роль. Внезапное изменение в поведении пользователя не всегда свидетельствует о подозрительной активности, поскольку определенные изменения могут быть связаны с выполнением дополнительных функций, ранее не входивших в круг обязанностей данного пользователя, но свойственных его роли, например при начале выполнения нового проекта. Однако при этом необходимо учитывать, что характер данных изменений должен совпадать с изменениями, свойственными группе пользователей, имеющих такую же функциональную роль. Очевидно, что данные изменения происходят не одновременно, а в течение некоторого достаточно длительного интервала времени.

Исходя из изложенного, можно определить следующие исходные данные, необходимые для построения профиля пользователя: число входов/выходов в информационную систему, число входов в систему со своего компьютера/число входов в систему с компьютера другого пользователя; число подключений съемных носителей информации к своему компьютеру/к компьютеру другого пользователя; число просмотренных файлов, число компьютеров, на которых были просмотрены файлы, число просмотренных файлов на своем компьютере/на компьютере другого пользователя; общее число отправленных/полученных электронных писем, число различных корпоративных и внешних получателей, а также минимальные, максимальные и средние значения этих параметров, вычисленные для группы пользователей, имеющих ту же роль в информационной системе, за заданный период времени.

При разработке моделей визуализации и проектировании графического интерфейса системы, поддерживающей визуальный анализ данных, следует придерживаться следующего принципа исследования данных, сформулированного Б. Шнайдерманом: «общий вид – выявление наиболее интересных паттернов – детали по требованию». В результате был разработан графический интерфейс программы (рис. 1).

Главное окно разделено на 4 окна: панель обзора статистических данных за неделю/месяц (A); статистические данные за выбранный интервал времени (сутки/неделю) (D); исходные данные за выбранный интервал времени (сутки/неделю) (C); облако слов (B).

Панель «интервал времени» принимается за отправную точку исследования данных. С его помощью можно выявить как характерные, так и не характерные для пользователя данной группы поведенческие изменения во времени. Выявив интервал времени, представляющий интерес для более пристального изучения, аналитик выбирает соответствующий глиф мышью, обновляет информацию, отображаемую на панелях B, C, D, переходят таким образом на требуемый уровень детализации исследуемых данных.

Основными элементами графического интерфейса пользователя являются модель визуализации «Часы» (окно D), линейный график с времен-

[2] OSSIM Website https://www.alienvault.com/open-threat-exchange/projects, 01.06.2015.

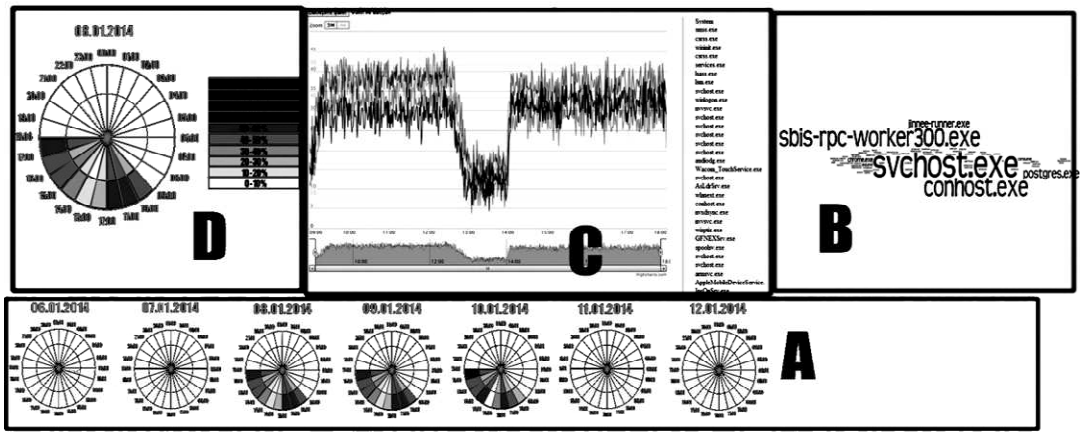


Рис. 1

ной шкалой (окно C) и облако слов (окно B). Модель визуализации данных «Часы» является адаптированной моделью графического представления данных, разработанной К. Кинтцелем и др. (C. Kintzel, J. Fuchs and F. Mansmann. Monitoring Large IP Spaces with ClockView). Ее достоинство – возможность компактно и естественным образом отобразить изменения параметра во времени. Для этого круг часов (рис. 2) разбивается на 24/7 равных сектора, соответствующих 24 ч в сутках/7 дням в неделе. Цветом обозначается среднее значение анализируемого параметра, рассчитанное за выбранный интервал времени. В данной статье эта модель используется для отображения набора параметров, имеющих одинаковый тип значений, например активность пользователя и загрузку ЦПУ, выраженные в процентах. Эта же модель применяется для описания статистического профиля пользователя за более длительный период времени (неделя, месяц). Такое решение позволяет определить закономерности в поведении пользователя, выявляя как периодически возникающие закономерности, так и «нестандартные» отклонения, требующие более пристального внимания со стороны аналитика. Те значения параметров, которые превышают статистические значения, вычисленные для группы пользователей, имеющих ту же роль, подсвечиваются цветом контура сектора.

Детально изучить изменение параметра за выбранный интервал времени можно при помощи линейного графика, на котором помимо текущего значения отображаются минимальное, среднее и максимальное значения параметра, вычисленные для группы пользователей с той же ролью. Таким образом, применение данного графика позволяет сузить исследуемый интервал времени.

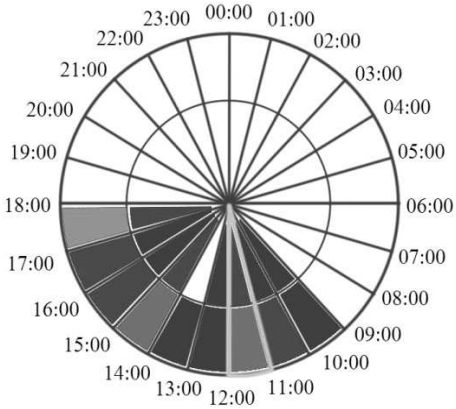


Рис. 2

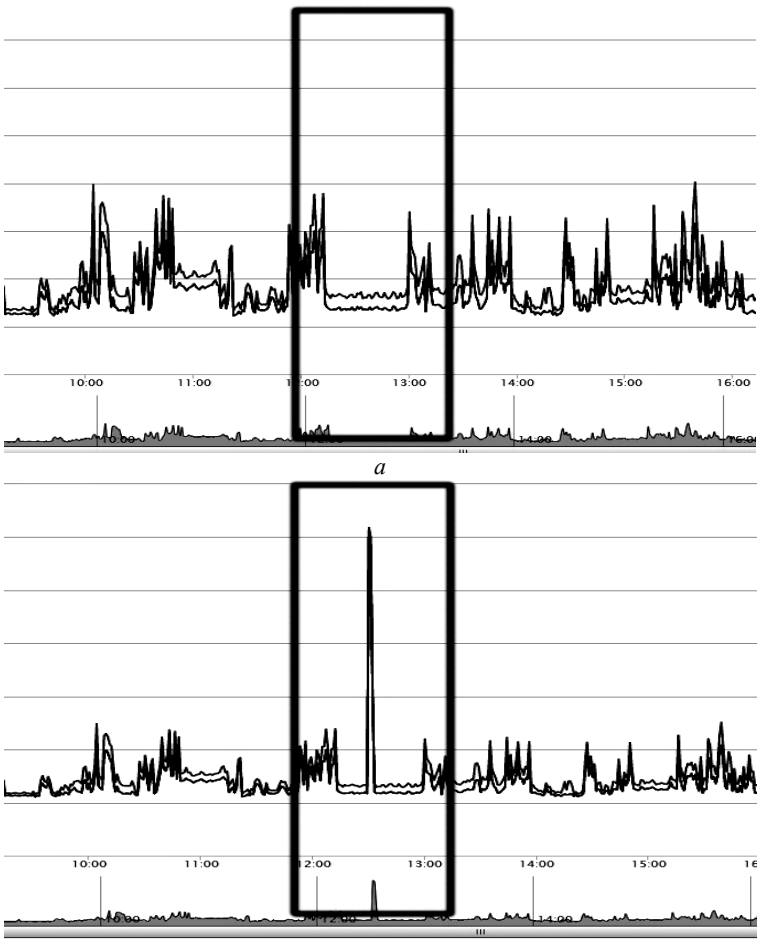
Для анализа текстовых данных используется методика визуального анализа текста – облако слов (тегов), которая позволяет быстро выявить наиболее часто используемые ключевые слова. Чем чаще встречается слово, тем больше размер его шрифта. Эта модель служит для определения наиболее часто используемых получателей электронных писем, процессов и приложений, названий файлов и съемных носителей. Цветом выделяются слова, которые не входят в множество наиболее часто встречаемых слов, характерных для данной группы пользователей.

С помощью разработанной методики были проведены эксперименты. Для оценки эффективности предложенного подхода был рассмотрен следующий сценарий внутреннего нарушителя. Злоумышленник в обеденный перерыв удаленно подключается к компьютеру пользователя с правами администратора, что позволяет ему получить полный доступ к информационным ресурсам системы, например для поиска и просмотра файловых данных. Исходные данные для эксперимента были сформированы добавлением искусственным образом логов, описывающих

нарушение безопасности, в реальные данные, собранные в течение месяца. Реальные исходные данные были получены от множества компьютеров, пользователи которых имели роль «разработчик программного обеспечения».

Собирались следующие данные: нагрузка ЦПУ, активность пользователя, список выполняемых процессов, открытие файлов, закрытие файлов, вход в систему и выход из нее. Предполагалось, что полученные данные не содержат аномалий, т. е. все сотрудники организации соблюдали распорядок рабочего времени и выполняли свои функциональные обязанности. Следовательно, можно ожидать, что определенные данные, например уровень нагрузки ЦПУ, активность пользователя, число нажатий на клавиатуру, должны возрастать в рабочее время, а во время обеда уменьшаться. На их основе был построен профиль пользователя во времени. Следует отметить, что для получения более точных результатов необходимо увеличить период сбора данных, потому что человеческий фактор играет критическую роль в сборе выбранных данных.

Дальнейшее изучение позволяет уточнить временной интервал возможной аномалии. На рис. 3, *а* представлены графики активности атакуемого пользователя без аномальной активности, а на рис. 3, *б* – с аномальной активностью. Рамкой выделена область, характеризующаяся низкой активностью пользователя, которая позволяет предположить, что сотрудник не находился в это время на рабочем месте. Данное время соответствует обеденному перерыву. Следует отметить, что в конце рабочего дня нагрузка ЦПУ увеличилась, но это, скорее, связанно с более активными действиями пользователя по выполнению поставленных задач. Данную активность можно объяснить отношением человека к работе. Например, кто-то более активен с утра, а к концу дня активность уменьшается, а кто-то, наоборот, под конец дня начинает чувствовать себя бодрее, следовательно, становиться более активным.



а

б

Рис. 3

На рис. 3, *б* выделена область, в которой зарегистрирована аномальная активность пользователя. В это же время был зарегистрирован сетевой вход в систему другого пользователя с правами администратора. Исследование других аспектов поведения позволяет аналитику сделать вывод о том, что злоумышленник искал определенную информацию, о чем свидетельствует увеличившееся число открытых файлов. Оценить характер искомой информации возможно, изучив формат открываемых файлов. В данном случае в перечне форматов просмотренных файлов наиболее часто встречались JPEG и PNG, что позволяет предположить, что злоумышленник искал информацию в файлах изображений, а также можно утверждать, что атакующий знал, где искать информацию, что позволит сузить круг подозреваемых.

Описанная методика отражает предложенный подход к обнаружению злоумышленника на локальном компьютере. Данный подход является актуальным, так как каждый день потенциальные злоумышленники улучшают свои навыки, улучшают стратегии и инструменты для совершения атак. Разработанная методика визуального анализа основана на формировании профиля положительного пользователя, что дает шанс выявить злоумышленника, который использует новую стратегию атаки, в то время как другое программное обеспечение по выявлению злоумышленника нацелено на стандартные, уже изученные стратегии.

В настоящее время продолжаются работы по улучшению модели визуализации, по усложнению сценария внутреннего нарушения информационной безопасности, а также по расширению собираемых данных, которые дадут более точный портрет пользователя, а следовательно, позволят уменьшить количество ложноположительных срабатываний. Также на стадии проектирования находится модуль автоматического анализа данных и модуль автоматического реагирования на атаку, для того чтобы система могла оповещать пользователя об инциденте безопасности и тот мог предотвратить действия злоумышленника или система могла сама отреагировать на вредоносные действия.

СПИСОК ЛИТЕРАТУРЫ

1. Salem M., Stolfo S. Modeling user search behavior for masquerade detection // Proc. of the 14th Intern. conf. on Recent Advances in Intrusion Detection, Menlo Park, CA, 2011. P. 181–200.

2. Greitzer F., Frincke D. Combining traditional computer security audit data with psychosocial data: predictive modeling for insider threat. // Insider Threats in Cyber Security. Advances in Information Security. 2010. Vol. 49. P. 85–114.

3. Cappelli D. M., Moore A., Trzeciak R. The CERT Guide to Insider Threats: How to Prevent, Detect and Respond to Information Technology Crimes (Theft, Sabotage, Fraud). N. Y.: Addison-Wesley Professional, 2012. 432 p.

4. Visual Analytics: Definition, Process, and Challenges / D. Keim, G. Andrienko, J.-D. Fekete, C. Goerg, J. Kohlhammer, G. Melancon // Information Visualization, Lecture Notes in Computer Science. 2008. Vol. 4950. P. 154–175.

5. Marty R. Applied Security Visualization. N. Y.: Addison Wesley Professional, 2008. 552 p.

6. Visual Analysis for Textual Relationships in Digital Forensics Evidence / T. Jankun-Kelly, D. Wilson, A. Stamps, J. Franck, J. Carver, J. Swan II // Proc. of 6th Intern. Workshop on Visualization for Cyber Security, Atlantic City, 11 Oct. 2009. P. 39–44.

E. S. Fedotov
Saint Petersburg state electrotechnical university «LETI»

DEVELOPMENT OF METHODS OF VISUAL ANALYSIS FOR FINDING OUT AN INTERNAL VIOLATOR

Devoted research of task of development of the method of visual analysis for finding out an internal violator. A research object is a process of exposure of anomalous activity in the conduct of user. The method of construction of visual type of user is examined in the article. Also the example of exposure of anomalous activity and its analysis is resulted in this work.

Internal ator, visual analysis, analysis of dens of the system, exposure of anomalies