Бегичева Светлана Викторовна,
старший преподаватель кафедры бизнес-информатики,
УрГЭУ,
г. Екатеринбург

ОБЗОР МЕТОДОВ ОБНАРУЖЕНИЯ СЕТЕВЫХ АНОМАЛИЙ

Аннотация. Методы обнаружения вторжений на основе детектирования аномалий применяются с целью заблаговременного реагирования и планирования действий по предотвращению вторжений. В статье приводится фундаментальная классификация подходов к обнаружению аномалий в данных с использованием методов интеллектуального анализа данных.

***Ключевые слова*:** поиск аномалий, интеллектуальный анализ данных, методы выявления аномалий

В течение последних лет вопросы информационной безопасности вызывают интерес как у производителей программного обеспечения, так и у обычных пользователей. Первые стремятся максимально защитить свой продукт от атак извне, чтобы не понести ущерб для репутации и свести к нулю материальные издержки, связанные с утечкой коммерчески важных данных. Вторые же считают важным защитить свои личные данные от несанкционированного доступа и распространения.

В связи с ростом вычислительной мощности и количества узлов в компьютерных сетях увеличивается и количество проходящих через них данных, что влечет за собой необходимость применения все более совершенных подходов к задаче обеспечения информационной безопасности. Важной стадией решения этой задачи является автоматизация сбора и анализа всего сетевого трафика, проходящего через узел при условии минимизации скорости обработки запросов для обеспечения разумных пределов задержки отклика.

До недавнего времени основным механизмом защиты корпоративных сетей от являлись межсетевые экраны, которые несмотря на то, что были предназначены для защиты информационных ресурсов, часто сами являлись уязвимыми. Это происходило из-за того, что администраторы сети были вынуждены упрощать систему контроля доступа для удобства пользователей, что значительно ослабляло защиту и создавало прорехи в безопасности ресурсов. К тому же межсетевые экраны увеличивают нагрузку на сеть, что существенно снижает скорость ее работы. Таким образом, использование межсетевых экранов не только не увеличивает безопасность корпоративной сети, но и снижает ее производительность, что может стать критическим фактором для высоконагруженных сетей, на которые накладывается требование высокого уровня быстродействия.

В случае с высоконагруженной сетью гораздо эффективнее сделать упор не на ужесточение контроля доступа, а на усовершенствование методов обнаружения вторжений в сеть и реагирования на них. Для круглосуточного мониторинга корпоративной сети для обнаружения атак предназначены так называемые системы активной защиты. Данные системы детектируют атаки на узлы корпоративной сети и реагируют на них заданным образом: например, заносят информацию об аномалии в регистрационные журналы и отправляют оповещение дежурному администратору.

В большинстве предметных областей нормальное поведение предопределено. Аномалия – это те шаблоны данных, которые не соответствуют норме, сильно отличаются и могут навредить или привести к разрушению системы [6].

Аномалии могут появляться в данных по разным причинам в разные моменты времени. Отдельно стоит отметить вредоносные причины появления аномалий: вторжение в систему безопасности, террористическая атака, техническая поломка и т.д. Эти причины имеют общие характеристики и ввиду своей опасности представляют большой интерес для исследователей.

Наиболее популярными на сегодняшний день методами детектирования аномалий являются приемы машинного обучения, которые с высокой точностью определяют случайные аберрации благодаря статистическим моделям.

В источниках встречается следующая классификация фундаментальных подходов к обнаружению аномалий в данных [5]:

- без данных об аномальности примера (кластеризация);
- классификация на основании данных о нормальном и аномальном поведении;
- классификация на основании данных только о нормальном или аномальном поведении.

Кластеризация, или определение выбросов без заранее размеченной выборки является примером обучения без учителя [3]. Данная концепция подразумевает, что точки множеств располагаются на основании статистического распределения, а самые отдаленные от соседей точки помечаются как потенциально аномальные. В таком случае считается что нормальные данные сосредоточены в конкретном месте, а аномалии лежат на некотором отдалении от них. Представленный класс методов подразумевает статичность и доступность полного набора данных для обработки.

Кластеризация выборки без учителя может происходить по двум сценариям:

- диагностика – выделение возможных аномалий и исключение их из дальнейшей обработки, что понижает вероятность переобучения модели;
- внедрение –умышленное использование потенциальных аберраций/ Таким образом вокруг нормальных данных помечается граница нормального поведения.

Обнаружение аномалий на основе создания модели нормального и аномального поведения [2] обязательно происходит с учителем, в роли которого выступает заранее подготовленная обучающая выборка, в которой размечены все записи. Важным условием является то, что модель должна быть обучена на данных, в полной мере соответствующих защищаемой системе или процессу, то есть

насколько полными будут изначальные познания модели о процессе, тем эффективней будет срабатывать классификатор.

Определение выбросов на основе модели только одной линии поведения, нормальной или аномальной [4] в ряде источников называют «поиск новинок» и относят к типу обучения частично с учителем [1]. Суть заключается в том, что обучение модели происходит только на одном из классов, чаще всего нормальном, так как собрать достаточной информации об аномалиях невозможно.

Такие модели способны распознавать аберрации как в данных динамического типа, так и статического. Детектирование происходит путём сверки нового примера с обучающим множеством. В случае, если наблюдение сравнимо с множеством, классификатор присваивает ему класс «норма», в противном случае «аномалия». Данная модель будет игнорировать все наблюдения, находящиеся за пределами границ нормального множества.

Определение аномалий является частным случаем задачи нахождения шаблонов данных, которые не соответствуют предсказанному состоянию [7]. Именно несоответствующие шаблоны являются аномалиями (или, в случае другой предметной области, всплесками, нестандартными наблюдениями, исключениями, аберрациями, особенностями). Поиск и выявление аномалий широко используется в области компьютерной безопасности, системах обнаружения вторжений, в диагностической медицине, транспортных системах, военном деле и т.д. Крайне важно вовремя обнаруживать аномалии и предупреждать их глобальное появление.

Список литературы

1. Chapelle, O. Semi-supervised learning / O. Chapelle, B. Schölkopf, A. Zien – Cambridge: MIT press, 2006. – Vol. 2.

2. Cunningham, P. Supervised Learning / P. Cunningham, M. Cord, S. J. Delany // Machine Learning Techniques for Multimedia. – Springer Berlin Heidelberg, 2008. – pp. 21-49.

3. Ghahramani, Z. Unsupervised learning / Z Ghahramani // Advanced Lectures on Machine Learning. – Springer Berlin Heidelberg, 2004. – pp. 72-112.

4. Khan, S. S. A survey of recent trends in one class classification / S. S. Khan, M. G. Madden // Artificial Intelligence and Cognitive Science. – Springer Berlin Heidelberg, 2010. – pp. 188-197.

5. Суханов, А.В. Интеллектуальные методы обнаружения и прогнозирования аномальных событий в темпоральных данных / А.В. Суханов // Диссертация на соискание ученой степени кандидата технических наук – Ростов-на-Дону, 2015. – С. 12-31.

6. Суханов, А.В. Метод нахождения аномалий при диагностике верхнего строения пути / А.В. Суханов, С.М. Ковалев// Программные системы и вычислительные методы — № 2(3) – Москва, NOTA BENE (ООО "НБ-Медиа"), 2013. – С. 176-180.

7. Drucker H., Wu D., Vapnik V. N. Support vector machines for spam categorization // IEEE Transactions on Neural networks. – 1999. – Vol. 10. – №. 5. – pp. 1048-1054.