щаемых сетей. Такой подход к обнаружению атак не требует длительных и сложных настроек. Администратору достаточно воссоздать в ВИС защищаемую сеть настолько, насколько это позволит текущий набор объектов в ВИС, периодически обновлять объекты. Несложно создать методику автоматической оценки системой защищаемой сети и создания ее аналога в ВИС.

Достоинство данного метода в том, что потенциально возможностей аномалий при обработке трафика гораздо меньше, чем возможное количество сигнатур атак. При этом методы обнаружения аномалий при обработке трафика для различных объектов часто будут достаточно сходны, что еще уменьшает затраты времени на разработку.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Абрамов Е.С.* Разработка комбинированной архитектуры системы обнаружения и выявления сетевых атак. Тезисы докладов на VII Всероссийской НК студентов и аспирантов «Техническая кибернетика, радиоэлектроника и системы управления». – Таганрог: Изд-во ТРТУ, 2004. – 560 с.

2. *Абрамов Е.С., Мордвин Д.В.* Разработка инструментальных средств для моделирования ЛВС // Материалы IX международной научно-практической конференции «Информационная безопасность». – Таганрог: Изд-во ТТИ ЮФУ, 2007.

3. *Абрамов Е.С., Мордвин Д.В.* Создание ложных информационных объектов для противодействия атакам в ЛВС. Материалы X международной научно-практической конференции «Информационная безопасность». – Таганрог: Изд-во ТТИ ЮФУ, 2008.

УДК 004.056

В.А. Нестеренко

ПОСТРОЕНИЕ И ИСПОЛЬЗОВАНИЕ ФУНКЦИИ ПЛОТНОСТИ В ПРОСТРАНСТВЕ ХАРАКТЕРИСТИК ДЛЯ ВЫЯВЛЕНИЯ АНОМАЛЬНЫХ СОБЫТИЙ

Современные методы обнаружения вторжений, основанные на выявлении аномальных событий, в основном построены по следующему принципу: вначале создаётся модель нормального поведения, характеризующая состояние системы в отсутствие нарушений, а затем выявляются отклонения наблюдаемых данных от нормального поведения [1]. Одним из методов построения модели нормального поведения системы является кластеризация данных [2]. В основе этого метода лежит допущение о том, что в пространстве числовых характеристик событий множество данных кластеризуется – собирается в отдельные группы. При подходящем выборе набора характеристик нормальные и аномальные события различимы – они группируются в разных кластерах. Кроме того, делается предположение о том, что число нормальных событий заметно превышает число аномальных событий, т.е. нормальные события формируют большие кластеры.

Для кластеризации данных применяются различные методы [3]: разделяющие методы, решёточные методы, методы основанные на плотности. В предлагаемой работе рассматривается метод кластеризации использующий оценку плотности в пространстве характеристик событий. Основная идея этого метода заключается в том, что некоторая область пространства принадлежит кластеру в том случае, если плотность событий в этой области превышает заданную пороговую величину. Подобные методы хорошо подходят для фильтрации выбросов и для создания кластеров произвольной формы.

Обозначим через D_1, D_2,..., D_N элементы входного набора данных и через D пространство всех возможных элементов данных или пространство событий. Конкретный вид пространства D зависит от вида анализируемых данных (приходящие пакеты, сетевые соединений, последовательность системных вызовов и т.п.). Для численного анализа данных отобразим элементы пространства событий в пространство характеристик X. Размерность пространства характеристик определяется количеством используемых признаков события D_i. В данной работе мы используем двумерное пространство характеристик: каждому событию D_i соответствуют две числовые, соответствующим образом нормированные, характеристики x_i и y_i. Предлагаемый метод может быть легко обобщён для случая пространства характеристик другой размерности. В качестве метрики и для оценки степени близости событий D_i и D_j будем использовать евклидово расстояние в пространстве характеристик:

$$d\left(D_i, D_j\right) = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2}.$$

Введём отношение порядка на множестве событий D: будем считать, что между парами событий D_1 и D_2 существует отношение $D_1 \le D_2$, если для соответствующих характеристик $X_1 = (x_1, y_1)$ и $X_2 = (x_2, y_2)$ одновременно выполняются условия $x_1 \le x_2$ и $y_1 \le y_2$. Другими словами, если в двумерном пространстве характеристик точка (x_2, y_2) расположена в первом квадранте относительно точки (x_1, y_1), то $D_1 \le D_2$.

Для хранения, обработки и изменения информации о текущих событиях мы будем использовать бинарное упорядоченное дерево событий T. Каждому событию D_i в дереве событий соответствует узел $node_i$, в этом узле хранится следующая информация: x_i и y_i – характеристики данного события и n_i^R – количество элементов в правом поддереве T_i^R узла $node_i$. Таким образом, бинарное дерево событий можно представить в виде

$$T = tree\left(node, T^L, T^R\right),$$

где $node = \left\{x, y, n^R\right\}$ – информация о событии, соответствующем данному узлу дерева, T^L и T^R – левое и правое поддеревья для данного узла.

Так как мы используем упорядоченное дерево, то для каждого элемента правого поддерева выполняется соотношение: $D_i \le D_j$, $D_j \in T_i^R$. В этом случае применимы стандартные алгоритмы включения новых и поиск заданных элементов в бинарном упорядоченном дереве.

Введённые отношения порядка и упорядоченность дерева событий позволяют находить количество элементов дерева T расположенных в первом квадранте относительно заданной точки (x_0, y_0). Для этой цели определим функцию $count\left(T, x_0, y_0\right)$, которая в качестве своего значения возвращает искомое число эле-

ментов (определение этой функции дано в приложении). В этом случае плотность данных в δ-окрестности точки (x_0, y_0) может быть вычислена по формуле

$$\rho(x_0, y_0) = \frac{count(T, x_1, y_1) - count(T, x_2, y_1) - count(T, x_1, y_2) + count(T, x_2, y_2)}{4 \cdot \delta^2},$$

где $x_1 = x_0 - \delta$, $x_2 = x_0 + \delta$, $y_1 = y_0 - \delta$, $y_2 = y_0 + \delta$.

Найденная функция плотности в пространстве характеристик событий может быть использована для выявления аномальных событий. Большие значения функции плотности выделяет области кластеров нормальных событий. Если очередное событие D_i попадает в такую область, то его можно считать нормальным, в противном случае – аномальным.

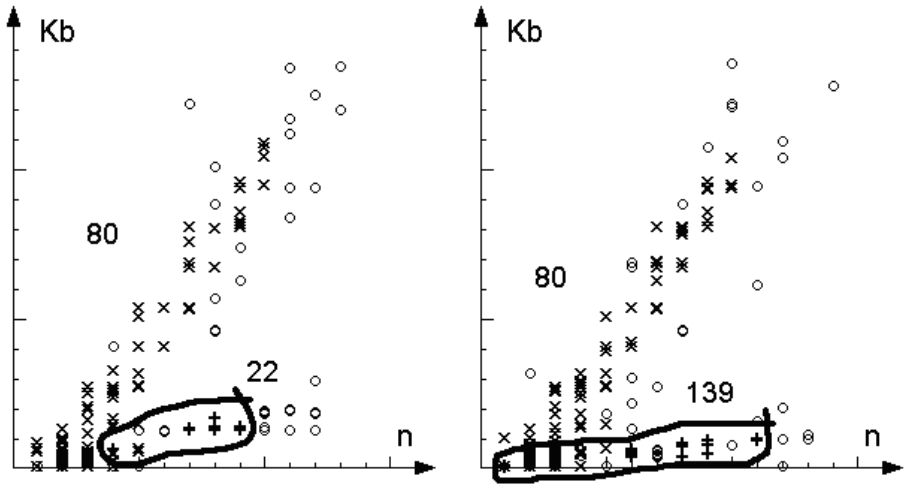
В качестве примера практического использования предлагаемого метода приведём результаты обработки потока пакетов в сети на одном из серверов факультета математики, механики и компьютерных наук ЮФУ. Log-файл сетевого трафика содержит информацию о более чем 1 млн пакетов и 200 тыс. корректных TCP соединениях.

Событиями в сети будем считать нормально установленные и правильно завершённые TCP соединения. В качестве числовых характеристик события будем использовать количество пакетов и объём информации переданных в течение одного соединения со стороны сервера – узла приёмника соединения.

На рис. 1 представлено распределение событий в двумерном пространстве характеристик. Точки изображённые крестиками соответствуют области высокой плотности: $\rho(x, y) \geq \rho_0$, где ρ_0 – средняя плотность в пространстве событий; светлые кружки – область низкой плотности. На каждом рисунке присутствует около 5 000 событий. Многие точки, представляющие отдельные события, расположены очень близко друг к другу, на рисунках они не различимы.

События представленные на рис. 1, а соответствуют соединениям со службами 80-го порта (http протокол) и 22-го порта (ssh – Security Shell Login). Кластер, соответствующий соединениям с 80-м портом (область этого кластера обозначена косыми крестиками), имеет вытянутую форму и расположен вдоль воображаемой линии проходящей через начало координат. Такая форма кластера объясняется тем, что протокол http используется в основном для передачи данных (содержимого Web-страницы) и объём переданной информации пропорционален количеству пакетов. Выделенная на рис. 1, а область соответствует соединениям с 22-м портом (точки этого кластера обозначены прямыми крестиками). Этот кластер довольно компактен и относительно изолирован от кластера 80-го порта. Хорошо видна разница в форме двух кластеров разных портов (областей высокой плотности) на рисунке.

Точки на рис. 1, б соответствуют соединениям со службами 80-го порта и 139-го порта (netbios сервис). Выделенная область соответствует соединениям с 139-м портом (точки этого кластера обозначены прямыми крестиками). Этот кластер расположен вдоль горизонтальной оси и, в отличие от предыдущего случая, сильно перекрывается с кластером 80-го порта. Если не учитывать соединения с малым числом переданных пакетов, то кластеры 80-го и 139-го портов также можно разделить.



а *б*

Рис. 1. Распределение событий в двумерном пространстве характеристик

Указанные свойства кластеров могут быть использованы для выявления туннелирования других протоколов через http-соединение, так как в этом случае изменится характерная форма кластера соответствующего соединениям с 80-м портом. Кроме того, предлагаемые методы построения функции плотности могут быть использованы для выявления аномальных событий на базе классификационной модели системы [4]: вначале определяем принадлежность события тому или иному кластеру (если это возможно), а затем определяем соответствие номера порта источника пакета с выбранным кластером.

Приложение. Подсчёт числа точек в пространстве событий.

В этом разделе мы рассмотрим реализацию используемой ранее операции подсчёта элементов дерева $count(T, x, y)$. Для наглядности при изображении данных в пространстве характеристик, для каждой точки в пространстве характеристик мы будем изображать сопутствующий 1-й квадрант.

Введённое отношение порядка $D_1 \le D_2$ (в пространстве характеристик событие D_2 расположено в первом квадранте относительно события D_1 – рис. 2, а) не может быть установлено для всех пар объектов из множества D (пример на рис. 2, б). Это обстоятельство требует дополнительного рассмотрения некоторых операций над элементами дерева (исключение узлов, объединение поддеревьев и т.п.).

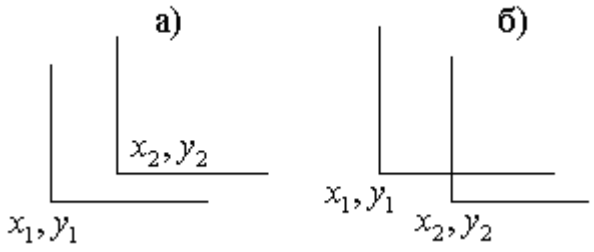


Рис. 2. Подсчёт числа точек в пространстве событий

Для нахождения плотности событий $\rho(x_0, y_0)$ в пространстве характеристик вблизи заданной точки (x_0, y_0) введём функцию $count(T, x_0, y_0)$. Эта функция будет использовать информацию хранящуюся в узлах дерева T и в качестве своего значения будет возвращать количество точек расположенных в первом квадранте относительно заданной точки (x_0, y_0). Обозначим через D_T событие соответствующее вершине дерева T, а через D_0 событие соответствующее заданной точке (x_0, y_0). Функция $count(T, x_0, y_0)$ должна подсчитывать число событий удовлетворяющих условию $D_i \geq D_0$, $1 \leq i \leq N$. Возможные значения функции $count(T, x_0, y_0)$ зависят от отношений порядка между событиями D_T и D_0.

1. В случае пустого дерева $T = nil$ значение функции $count(T, x_0, y_0)$ очевидно равно нулю.

2. Если $D_T \leq D_0$, то ни один из нужных нам элементов не будет находиться в левом поддереве T^L дерева T и, следовательно, $count(T, x_0, y_0) = count(T^R, x_0, y_0)$.

3. Если $D_T \geq D_0$, то в число требуемых элементы входят все элементы правого поддерева T^R и частично элементы левого поддерева T^L, но информация о количестве элементов в правом поддереве n^R содержится в корневом узле дерева T и поэтому $count(T, x_0, y_0) = n^R + count(T^L, x_0, y_0)$.

4. В случае если между событиями D_T и D_0 нет отношений порядка, то $count(T, x_0, y_0) = count(T^L, x_0, y_0) + count(T^R, x_0, y_0)$.

Учитывая все рассмотренные варианты приходим к следующему определению функции $count(T, x_0, y_0)$:

$$count(T, x_0, y_0) = \begin{cases} 0 \quad \text{если} \quad T = \text{nil} \\ count(T^R, x_0, y_0) \quad \text{если} \quad x \leq x_0 \wedge y \leq y_0 \\ n^R + count(T^L, x_0, y_0) \quad \text{если} \quad x > x_0 \wedge y > y_0 \\ count(T^L, x_0, y_0) + count(T^R, x_0, y_0) \end{cases}$$

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *D.E. Denning*. An intrusion detection model. IEEE Transactions on Software Engineering, SE-13. 1987. P. 222-232.

2. *L. Portnoy, E. Eskin and S. J. Stolfo*. Intrusion Detection with Unlabeled Data Using Clustering. In Proceedings of ACM CSS Workshop on Data Mining Applied to Security (DMSA-2001), Philadelphia, PA, 2001.

3. *E. Eskin, A. Arnold, M. Prerau, L. Portnoy, and S. Stolfo*. A geometric framework for unsupervised anomaly detection: detecting intrusions in unlabeled data Applications of Data Mining in Computer Security, Kluwer, 2002.

4. *Wenke Lee and Sal Stolfo*. Data Mining Approaches for Intrusion Detection. In Proceedings of the 7th USENIX Security Symposium (SECURITY'98), San Antonio, Texas, January 26-29, 1998.