УДК 004.056, 519.812.3

А. В. Гирик

ПРИМЕНЕНИЕ МЕТОДОВ МНОГОКРИТЕРИАЛЬНОГО ПРОГНОЗИРОВАНИЯ В СЕТЕВЫХ СИСТЕМАХ ОБНАРУЖЕНИЯ ВТОРЖЕНИЙ

Рассматриваются вопросы разработки математических моделей для сетевых систем обнаружения вторжений. Предложен метод обнаружения сетевых аномалий на основе многокритериального прогнозирования показателей безопасности. Поставлена и решена задача расчета весовых коэффициентов, учитывающих значимость моделей прогнозирования, которая может изменяться с течением времени. Предложенный подход обладает малой вычислительной сложностью и не зависит от характера критериев выбора прогнозов.

Ключевые слова: *обнаружение вторжений, обнаружение сетевых аномалий, сетевая безопасность, многокритериальное прогнозирование.*

Широкое распространение разнообразных информационных систем и расширение набора услуг, предоставляемых клиентам этих систем, обусловили увеличение количества информационных угроз безопасности сетей передачи данных. В связи с этим обеспечение своевременного выявления и идентификации угроз нарушения информационной безопасности является актуальной и важной проблемой. Как правило, она решается с помощью специального программного комплекса — распределенной системы обнаружения вторжений (Intrusion Detection System — IDS), которая осуществляет [1] мониторинг потоков данных в сети, анализ этих данных и выявление в анализируемых потоках данных признаков информационных угроз безопасности.

По принципу функционирования системы обнаружения вторжений делятся на два класса — выполняющие сигнатурный анализ и анализирующие статистику показателей безопасности. Особый интерес представляет обнаружение угроз безопасности сети в целом, так как последствия перегрузок, распределенных DoS-атак и других угроз, направленных на дестаби-

лизацию функционирования сети, приводят к значительным материальным потерям [1]. В общем случае угрозой можно считать аномальное поведение любого из наблюдаемых показателей безопасности. На практике показатели объединяют в группы и считают аномалию достоверно обнаруженной в случае, если поведение каждого показателя в группе идентифицировано как отклонение от нормального профиля сверх некоторого предельного значения.

Метод обнаружения сетевых аномалий в режиме реального времени заключается в следующем: выполняется мониторинг некоторого показателя безопасности, затем на основе накопленных данных строится прогноз, т.е. рассчитывается, какие значения показатель примет в ближайшее время, после чего прогноз сравнивается с реальными значениями показателя и на основании определенных критериев (в первую очередь, величины ошибки прогноза) принимается решение о наличии аномалии.

Эффективность метода зависит от точности прогноза. При решении задач обнаружения сетевых аномалий (в отличие, например, от задач планирования инфраструктуры) существует потребность в получении точных прогнозов с относительно небольшим горизонтом [2]. Для повышения точности прогноза предлагается одновременно использовать несколько моделей прогнозирования. В зависимости от времени и других факторов степень корректности моделей может изменяться, поэтому при формировании обобщенного прогноза необходимо определить степень значимости модели. Таким образом, приходим к задаче выбора весовых коэффициентов, учитывающих значимость модели, т.е. определение вклада, который вносит та или иная модель прогнозирования (тот или иной ее компонент) в формирование прогноза на очередном шаге.

Сформулированные задачи будем решать как задачи многокритериальной оптимизации [3]. Пусть имеется совокупность моделей прогнозирования Z_k, $k=1, 2, ..., N$, которые обеспечивают формирование прогнозов с некоторым горизонтом h (т.е. на h отсчетов вперед). Степень соответствия совокупности прогнозов X_j, $j=1, 2, ..., N$, реальным значениям параметра X определяется с помощью семейства критериев $Q=\{Q_1, Q_2, ..., Q_S\}$, которое содержит S частных критериев. Будем считать, что каждый критерий $Q_i(X_j)\in[0,1]$, иными словами, значение каждого частного критерия есть величина из интервала [0,1], причем „0“ означает полное несовпадение по данному критерию, а „1“ соответствует полному совпадению. Таким образом, частные критерии нормализованы и приведены к безразмерному виду. В этом случае нужно обеспечить $Q_i(X_j)\to\max$ для всех j.

Как правило, многокритериальная задача сводится к однокритериальной путем введения обобщенного критерия оптимальности F, который может быть аддитивной, мультипликативной или среднестепенной функцией частных критериев [4]. В простейшем случае можно считать, что все критерии обладают одинаковой значимостью. На практике, тем не менее, часто оказывается так, что в зависимости от времени и от исходных данных степень значимости критериев может изменяться. В этом случае каждому критерию ставится в соответствие весовой коэффициент, отражающий его значимость. Для определения значений коэффициентов используются различные методы, например диалоговый метод задания коэффициентов, в общем случае предполагающий, что лицо, принимающее решение, предоставит достаточные для расчета коэффициентов сведения [5]. Другой подход заключается в вычислении значений коэффициентов исходя из качественных характеристик критериев, например чувствительности. Будем считать, что частные критерии качественно соизмеримы между собой.

Введем следующие обозначения: $Q_i^{\text{opt}}=\max\limits_j Q_i(X_j)$ — оптимальное значение i-го критерия для совокупности прогнозов X_j, $j=1, 2, ..., N$; $X_i^{\text{opt}}=\arg\max\limits_j Q_i(X_j)$ — решение (прогноз), оптимальное по i-му критерию. Рассмотрим меру

$$C_{il} = \left| \frac{Q_i^{\text{opt}} - Q_i(X_l^{\text{opt}})}{Q_i^{\text{opt}}} \right|,$$

которая определяет относительное отклонение оптимального значения i-го критерия от значения i-го критерия, полученного для оптимального решения по l-му критерию, $i, l = 1, 2, ..., S$. Для всех возможных значений i и l построим матрицу из элементов C_{il}.

По значениям C_{il} можно судить о том, насколько чувствительным к решениям X_j является i-й критерий: для этого нужно проанализировать соответствующий столбец матрицы. Вычислим разность между максимальным и минимальным элементами всех столбцов и сформируем вектор $\mathbf{V}^T = (v_1, v_2, ..., v_S)$, где $v_i = \max\limits_l C_{il} - \min\limits_l C_{il}$. Весовые коэффициенты определяются как вектор $\mathbf{W}^T = (w_1, w_2, ..., w_S)$, где

$$w_k = v_i \Big/ \left(\sum_{t=1}^{S} v_t \right).$$

Таким образом, определим обобщенный критерий оптимальности для решения X_j в аддитивной форме:

$$F(X_j) = \sum_{i=1}^{S} w_i Q_i(X_j), \qquad (1)$$

тогда оптимальное решение будет найдено как $X^{\text{opt}} = \arg\max\limits_j F(X_j)$.

Если с течением времени предпочтения, определяющие выбор моделей, изменяются, то весовые коэффициенты для моделей прогнозирования могут быть рассчитаны с учетом динамики изменения значений обобщенного критерия. Рассчитаем эти значения для каждой из N моделей и последних M прогнозов. В результате получим матрицу $\mathbf{F}$, состоящую из элементов F_{kj}, где F_{kj} — значение обобщенного критерия, рассчитанного с использованием формулы (1) для k-й модели прогнозирования и прогноза X_{kj}, $j = 1, ..., M$, $k = 1, ..., N$. Рассчитаем весовые коэффициенты u_k для моделей следующим образом. Пусть $\alpha \in [0,1]$ — некоторая константа, $p_k = \sum\limits_{j=1}^{M} \alpha^{M-j+1} F_{ij}$, тогда

$$u_k = p_k \Big/ \left(\sum_{l=1}^{N} p_l \right).$$

Матрица $\mathbf{F}$ переформировывается по мере получения реальных значений параметров таким образом, чтобы значения коэффициентов отражали относительный вклад той или иной модели на очередном шаге. Если $Z_k(t)$ — k-я модель прогнозирования, то обобщенная модель процесса может быть представлена в виде

$$X(t) = \sum_{k=1}^{N} u_k Z_k(t). \qquad (2)$$

Для расчета процесса с помощью выражения (2) необходимо заполнить матрицу $\mathbf{F}$, т.е. нужно, чтобы модели были настроены на обучающей выборке, составляющей, по меньшей мере, M прогнозов.

Рассмотренный в настоящей статье подход, основанный на сравнении оперативных прогнозов показателей безопасности с их нормальным профилем, позволяет обеспечить повы-

шение точности обнаружения сетевых аномалий. Необходимо также отметить, что при защите сети следует ориентироваться не только на известные в настоящее время, но и на возможные угрозы, и, кроме того, учитывать особенности поведения показателей безопасности в конкретном сетевом окружении.

СПИСОК ЛИТЕРАТУРЫ

1. *Лукацкий А. В.* Системы обнаружения атак // Сетевой. 2002. № 4. С. 45—48.

2. *Будько М. Б., Будько М. Ю., Гирик А. В.* Применение авторегрессионного интегрированного скользящего среднего в алгоритмах управления перегрузками протоколов передачи потоковых данных // Науч.-техн. вестн. СПбГУ ИТМО. 2007. № 39. С. 319—323.

3. *Рыков А. С.* Методы системного анализа: многокритериальная и нечеткая оптимизация, моделирование и экспертные оценки. М.: Экономика, 1999. 191 с.

4. *Гайдышев И.* Анализ и обработка данных: Спец. справочник. СПб.: Питер, 2001. 752 с.

5. *Жигулин Г. П., Серебров А. И., Яковлев А. Д.* Прогнозирование устойчивости и функционирования объектов с использованием теории игр и исследования операций. СПб.: СПбГУ ИТМО, 2004. 204 с.

Сведения об авторе

Алексей Валерьевич Гирик — аспирант; Санкт-Петербургский государственный университет информационных технологий, механики и оптики, кафедра мониторинга и прогнозирования чрезвычайных ситуаций; E-mail: alexei.guirik@googlemail.com

Рекомендована кафедрой мониторинга и прогнозирования чрезвычайных ситуаций

Поступила в редакцию 25.04.08 г.