УДК 004.02

МЕТОД ОБНАРУЖЕНИЯ АНОМАЛИЙ В ПОТОКАХ ДАННЫХ СЕТЕВОГО ТРАФИКА

М.Ю. МАКАРЫЧЕВ, В.А. МАКАРЫЧЕВА
(Представлено: канд. техн. наук, доц. А.Ф. ОСЬКИН)

В данной статье рассматривается метод обнаружения сетевых аномалий с применением статистических алгоритмов над вейвлет-коэффициентами (аппроксимаций и детализаций), полученных путем вейвлет-анализа из последовательности коэффициентов сетевого трафика.

Введение. Характеристики трафика, проходящего по магистральным сетям связи, содержат в себе информацию о состоянии сетевого оборудования и подсетей, примыкающих к основной магистрали. Аномалии, зафиксированные в этом трафике, подлежат обязательной обработке. Они могут свидетельствовать о сетевых атаках, таких как DDoS (Distributed Denial of Service), отсутствии связи в некоторых сегментах сети, неисправности оборудования и других неполадках. Таким образом, своевременное обнаружение аномалии может гарантировать стабильную работу сети. В статье рассматривается метод обнаружения аномалий сетевого трафика, относящийся к методам кратномасштабного анализа.

Основной раздел. Для обнаружения аномалий предлагается использовать вейвлет-преобразование сигнальной кривой, отображающей зависимость «трафик-время». Одним из преимуществ вейвлет-преобразования является то, что оно дает возможность проанализировать сигнал в частотно-временной области и позволяет исследовать аномальный процесс на фоне остальных компонент [1].

Рассмотрим обнаружение аномалий сетевого трафика на основе дискретного вейвлет-преобразования с применением статистических критериев. Для адаптации этого способа к анализу трафика в реальном времени используется техника трех скользящих окон W_1, W_2, W_3. Первое окно является окном сравнения, второе и третье – окнами обнаружения. Пусть размер каждого окна будет ω_1, ω_2 и ω_3 выбранных временных единиц соответственно, причем $\omega_1 > \omega_2 > \omega_3$. Тогда в произвольный момент времени t начало окна W_3 будет находиться в точке t, в нем будут содержаться ω_3 значений трафика от $t-\omega_3$ до t, в окне W_2 – ω_2 значений от $t-\omega_3-\omega_2$ до $t-\omega_3$, а в окне W_1 – ω_1 значений от $t-\omega_3-\omega_2-\omega_1$ до $t-\omega_3-\omega_2$.

Выполнив быстрое вейвлет-преобразование для выборок внутри каждого из окон в каждый момент времени t_i, будет вычисляться на некотором масштабном уровне j набор коэффициентов для окна W_1 – аппроксимации $\{a_{1x}, a_{2x}, \ldots, a_{nx}\}_{t,j}$ и детализации $\{d_{1x}, d_{2x}, \ldots, d_{nx}\}_{t,j}$, для окна W_2 – аппроксимации $\{a_{1y}, a_{2y}, \ldots, a_{my}\}_{t,j}$ и детализации $\{d_{1y}, d_{2y}, \ldots, d_{my}\}_{t,j}$ и для окна W_3 – аппроксимации $\{a_{1z}, a_{2z}, \ldots, a_{kz}\}_{t,j}$ и детализации $\{d_{1z}, d_{2z}, \ldots, d_{kz}\}_{t,j}$. Причем количество коэффициентов n на уровне j в окне W_1 будет определяться выражением $n = \frac{\omega_1}{2^j}$, в окне W_2 – выражением $m = \frac{\omega_2}{2^j}$, а в окне W_3 – выражением $k = \frac{\omega_3}{2^j}$. Эти коэффициенты будут проверяться по статическим критериям, и на основе принятия или отклонения статистических гипотез будет выноситься решение о кардинальном различии в анализируемых параметрах между окнами W_1, W_2 и W_3, а следовательно, о наличии аномалии или же наоборот – их отсутствии.

Анализ статистических характеристик коэффициентов аппроксимации и детализации показывает, что плотность распределения вероятностей мгновенных значений этих коэффициентов хорошо описывается нормальным распределением [2]. Для обнаружения аномалий, выражающихся в изменении дисперсии, предлагается использовать критерий Бартлетта, а для обнаружения величины среднего значения – критерий Кохрена-Кокса.

Критерий Бартлетта предложен для обнаружения изменений в дисперсиях выборок окон W_1, W_2 и W_3. В каждый момент времени (положении окон) t на масштабном уровне j выдвигаются две статистические гипотезы о равенстве дисперсий трех выборок $\{d_{1x}, d_{2x}, \ldots, d_{nx}\}_{t,j}$, $\{d_{1y}, d_{2y}, \ldots, d_{my}\}_{t,j}$ и $\{d_{1z}, d_{2z}, \ldots, d_{kz}\}_{t,j}$: нулевая H_0: $\sigma^2_{1,t,j} = \sigma^2_{2,t,j} = \sigma^2_{3,t,j}$ и альтернативная H_1: $\sigma^2_{1,t,j} \neq \sigma^2_{2,t,j} \neq \sigma^2_{3,t,j}$.

Алгоритм обнаружения выбросов на основе анализа аномального изменения дисперсий записывается как $\chi^2_{t,j} = M_{t,j}\left[1+\frac{1}{6}\left(\frac{1}{n-1}+\frac{1}{m-1}+\frac{1}{k-1}\right)-\frac{1}{N}\right]^{-1}$. Введем обозначения:

$$M_{t,j} = N\ln\left[\frac{1}{N}\left((n-1)S^2_{1,t,j}+(m-1)S^2_{2,t,j}+(k-1)S^2_{3,t,j}\right)\right]-$$

$$-\left[(n-1)\ln S^2_{1,t,j}+(m-1)\ln S^2_{2,t,j}+(k-1)\ln S^2_{3,t,j}\right]$$

$$N = n+m+k-3$$

$S^2_{1,t,j} = \frac{1}{n-1}\sum_{i=1}^{n}\left(d_{ix}-\overline{d_x}\right)^2$ – выборочная дисперсия выборки последовательности деталей на масштабном уровне j в окне W_1.

$S^2_{2,t,j} = \frac{1}{m-1}\sum_{i=1}^{m}\left(d_{iy}-\overline{d_y}\right)^2$ – выборочная дисперсия выборки последовательности деталей на масштабном уровне j в окне W_2.

$S^2_{3,t,j} = \frac{1}{k-1}\sum_{i=1}^{k}\left(d_{iz}-\overline{d_z}\right)^2$ – выборочная дисперсия выборки последовательности деталей на масштабном уровне j в окне W_3.

$\overline{d_x} = \frac{1}{n}\sum_{i=1}^{n} d_{ix}$ – выборочное среднее выборки последовательности деталей на масштабном уровне j в окне W_1.

$\overline{d_y} = \frac{1}{m}\sum_{i=1}^{m} d_{iy}$ – выборочное среднее выборки последовательности деталей на масштабном уровне j в окне W_2.

$\overline{d_z} = \frac{1}{k}\sum_{i=1}^{k} d_{iz}$ – выборочное среднее выборки последовательности деталей на масштабном уровне j в окне W_3.

Нулевая гипотеза опровергается в пользу альтернативной, в случае если $\chi^2_{t,j} > \chi^2_{\alpha,2}$, где $\chi^2_{\alpha,2}$ – α-квантиль распределения хи-квадрат с двумя степенями свободы.

Для обнаружения изменений среднего значения выборок аппроксимаций $\{a_{1x}, a_{2x}, \ldots, a_{nx}\}_{t,j}$, $\{a_{1y}, a_{2y}, \ldots, a_{my}\}_{t,j}$ и $\{a_{1z}, a_{2z}, \ldots, a_{kz}\}_{t,j}$ предложен критерий Кохрена-Кокса. Данный критерий можно применять только для двух выборок, поэтому его можно применить только для окон W_1 и W_2, т.к. критерий предлагается в [2] для обнаружения долговременных низкочастотных аномалий. Далее будет представлен алгоритм обнаружения выбросов на основе анализа аномального изменения среднего значения выборки на примере окон W_1 и W_2.

Статистикой критерия является $Y = \frac{1}{S_{t,j}}\left(\frac{1}{n}\sum_{i=1}^{n} a_{ix} - \frac{1}{m}\sum_{i=1}^{m} a_{iy}\right)$. Введем, как и прежде, обозначения:

$S^2_{1,t,j} = \frac{1}{n-1}\sum_{i=1}^{n}\left(a_{ix}-\overline{a_x}\right)^2$ – выборочная дисперсия выборки последовательности аппроксимаций на масштабном уровне j в окне W_1.

$S^2_{2,t,j} = \frac{1}{m-1}\sum_{i=1}^{m}\left(a_{iy}-\overline{a_y}\right)^2$ – выборочная дисперсия выборки последовательности аппроксимаций на масштабном уровне j в окне W_2.

$S^2_{t,j} = \frac{S^2_{1,t,j}}{n}+\frac{S^2_{2,t,j}}{m}$ – суммарная взвешенная дисперсия выборок аппроксимаций для окон W_1 и W_2.

$\overline{a_x} = \frac{1}{n}\sum_{i=1}^{n} a_{ix}$ и $\overline{a_y} = \frac{1}{m}\sum_{i=1}^{m} a_{iy}$ – выборочные средние выборок последовательности аппроксимаций на масштабном уровне j в окне W_1 и W_2 соответственно.

С учетом введенных обозначений статистика сводится к виду $Y = \frac{1}{S_{t,j}}\left(\overline{a_y} - \overline{a_x}\right)$.

Критические (пороговые) значения статистики вычисляются по формуле: $t'_\alpha = \frac{f_1 t_\alpha(\upsilon_1) + f_2 t_\alpha(\upsilon_2)}{f_1 + f_2}$,

где $f_1 = \frac{S^2_{1,t,j}}{n}$, $f_2 = \frac{S^2_{2,t,j}}{m}$; $t_\alpha(\upsilon)$ – α-квантиль распределения Стьюдента с υ степенями свободы ($\upsilon_1 = n-1$ и $\upsilon_2 = m-1$).

Для каждого статистического критерия используются два порога, исходя из уровня значимости: $\alpha = 0.05$ и $\alpha = 0.01$. Превышение верхнего порога на каком-либо уровне вейвлет-разложения означает наличие аномалии. В случае превышения нижнего порога, производится дальнейшая декомпозиция по следующему уровню разложения, и для коэффициентов этого уровня опять будут проверяться статистические критерии.

Техника применения трех (а не двух) скользящих окон обоснована тем, что результат обнаружения зависит от размера окна обнаружения и длительности аномалии [3]. Размеры окон сравнения и обнаружения предполагается устанавливать экспериментально.

Заключение. Рассмотренный метод основывается на вейвлет-представлении временного ряда сетевого трафика и статистических алгоритмах обнаружения аномалий с двумя порогами. Метод может быть применен для класса задач обнаружения аномалий в компьютерных и телекоммуникационных сетях.

ЛИТЕРАТУРА

1. Максименко, Г.А. Метод обнаружения аномалий потоков данных в сетях / Г.А. Максименко // Системи обробки інформації: сб. науч. ст. / Харьковский Нац. ун-т Воздушных Сил им. И. Кожедуба; С.В. Кавун, Г.А. Кучук (отв. за выпуск) – Харьков, 2009. – выпуск 7 (81) – С. 33–37.
2. Шелухин, О.И. Обнаружение вторжений в компьютерные сети. Сетевые аномалии / О.И. Шелухин, Д.Ж. Сакалема, А.С. Филинова. – Горячая линия – телеком, 2013. – 220 с.
3. Шелухин, О.И. Сравнительный анализ характеристик обнаружения аномалий трафика методами кратномасштабного анализа / О.И. Шелухин, А.В. Панкрушин // T-Comm Телекоммуникации и транспорт. – 2014. – № 6. – С. 65–70.