УДК 621.396, 681.3

Г.А. Максименко

Национальная комиссия по регулированию связи Украины, Киев

МЕТОД ОБНАРУЖЕНИЯ АНОМАЛИЙ ПОТОКОВ ДАННЫХ В СЕТЯХ

Предлагается для решения задач обнаружения аномалий трафика в компьютерных и телекоммуникационных сетях метод, основанный на пакетном вейвлет-разложении данных трафика и статистическом алгоритме обнаружения использующем критерий Кохена. Критерий Кохена позволяет понизить уровень ложных тревог при обнаружении аномалий трафика на низких частотах. Данный алгоритм использует два скользящих окна для обнаружения аномалий, два пороговых значения и вследствие этого отличается от известных достаточно низким уровнем ложных тревог, высокой производительностью и адаптивностью.

Ключевые слова: *обнаружение, аномалия, трафик, вейвлет-преобразование.*

Введение

Типовая опорная телекоммуникационная сеть состоит из узлов соединенных меду собой линиями связи. Определим поток между исходящей точкой и конечным пунктом, как трафик, который входит в опорную сеть оператора связи в некотором исходном узле и выходит в узле, ближайшем к месту места назначения. Трафик наблюдаемой в каждой линии связи, представляет собой суперпозицию нескольких информационных потоков. Под термином **аномалия трафика** будем понимать, на определенном временном интервале, внезапные положительные или негативные изменения в потоке трафика. Трафик в маршрутизаторе магистральной телекоммуникационной сети меняется непрерывно.

Целью данной статьи является разработка метода обнаружения аномалий и своевременное выявление отклонений аномального трафика на фоне основного потока информации.

Основной материал

Предлагаемый метод является дальнейшим развитием работ, изложенных в [1 – 11] и позволяет обнаруживать и локализовывать самые незначительные изменения выборочных значений временного ряда.

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. Суть алгоритма вейвлет-декомпозиции состоит в

том, что расщепление компонент сигнала осуществляется не только в НЧ, но и в ВЧ области. При данном алгоритме операция расщепления или декомпозиции применяется к любой из получающихся ВЧ-компонент. В этом случае просто происходит замена вейвлета ш(t) на два новых вейвлета: $\psi_1(t)$ и $\psi_2(t)$. И так далее. Преобразование с помощью вейвлет-пакетов является **адаптивным**, поскольку оно легко приспосабливается к особенностям сигнала и может успешно использоваться не только для анализа сигналов, но и для их компрессии и очистки от шумов. Достоинством вейвлет-пакетов и адаптивных алгоритмов их реализации является отсутствие необходимости в обучении системы и даже в оценке статистических характеристик сигналов. Все, что нужно – это ввести оценку стоимости вейвлет-коэффициентов, мерой которой может служить энтропия – концентрация числа вейвлет-коэффициентов, требующихся для описания сигнала с некоторой заданной точностью. Далее, посредством адаптивной реконструкции вейвлет-коэффициентов различных вейвлет-областей содержащих признаки аномалий трафика, можно подтвердить параметры аномалий и увеличить надежность обнаружения. Благодаря методу пакетного вейвлет-преобразования с использованием скользящего окна можно уменьшить вычислительную сложность путем устранения избыточных вычислений. Путем применения окна и запоминания части коэффициентов в памяти можно избежать повторного избыточного вычисления, т.е. ускорить работу вычислительного алгоритма увеличив затраты на объем памяти.

Применение «скользящего окна» позволяет увеличить надежность обнаружения даже незначительных аномалий. Известно, что плотность спектральной мощности временного ряда «трафик – время», при наличии аномалий, имеет пики на определенных частотах. Вейвлет-анализ позволяет обнаружить аномалии трафика на основании различий спектров обычного и аномального трафика. В основу метода обнаружения положен модифицированный статистический алгоритм обнаружения аномалий – алгоритм среднеквадратического отклонения. Суть его состоит в том, что применяются два окна обнаружения для вычисления среднеквадратического отклонения данных временного ряда. Под данными понимается количество переданных пакетов или информации в единицу времени. Обозначим одно окно – как $\mathbf{W}_1$, а другое – окно обнаружения как $\mathbf{W}_2$. Оба окна перемещаются во времени для динамического обновления данных. В некоторый момент времени t, вычисляется среднеквадратическое отклонение временного ряда данных V_2 в окне обнаружения (t – $\mathbf{W}_2$, t) и среднеквадратическое отклонение временного ряда данных V_1 в более широком исходном окне $\mathbf{W}_1$, т.е. (t – $\mathbf{W}_1$, t). Определим отношение

$$Z = V_2 / V_1, \qquad (1)$$

где $V = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(f_i - \bar{f})^2}$; f_i – i-й элемент выборки; $\bar{f}$ – среднее арифметическое выборки.

Тогда параметр Z будет представлять собой величину стандартного отклонения выборки в окне обнаружения аномалий $\mathbf{W}_2$ по отношению к нормальным среднестатистическим колебаниям временного ряда в более широком окне $\mathbf{W}_1$. Если в окне обнаружения $\mathbf{W}_2$ появятся аномальные выбросы данных, тогда соответственно возрастет значение отношения Z. Данный алгоритм среднеквадратического отклонения количественного показателя Z в чистом виде используется после реконструкции сигнала, когда будет выполнена процедура избавления от шумов пороговым методом. Однако данный алгоритм предлагается использовать и на этапе первоначального обнаружения при разложении сигнального трафика на коэффициенты вейвлет-преобразования. В этом и состоит суть метода.

При первоначальном обнаружении аномалий в момент времени t, осуществляется пакетная вейвлет-декомпозиция временного ряда сетевого трафика на интервале (t – $\mathbf{W}_1$, t). В результате получаем набор вейвлет-коэффициентов при конкретных значениях масштаба, которые затем обрабатываются с помощью алгоритма среднеквадратического отклонения Z. По мере того, как величина отношения Z превысит некоторый пороговый уровень, то можно говорить о наличии аномалии в трафике. Однако полученные вейвлет-коэффициенты смешаны с шумами и могут иметь некоторые колебания значений. Для более четкого подтверждения наличия аномалии в трафике, осуществляется реконструкция сигнала по вейвлет-коэффициентам при тех же значениях масштаба и повторно применяется алгоритм обнаружения методом двух окон после выполненной при реконструкции процедуры избавления от шумов. Данный алгоритм достаточно эффективно определяет высокочастотные аномалии трафика, однако проблемой остается обнаружение низкочастотных аномалий.

Если использовать для обнаружения аномалий только алгоритм вычисления с.к.о., то тогда НЧ аномалии, длящиеся большой интервал времени, могут привести к ложным тревогам. Главная причина заключается в том, что низкочастотные аномалии слабо меняются или сохраняют стабильность при достижения определенного значения, кроме того если окно обнаружения значительно короче, чем длительность НЧ аномалии, то величина Z, представляющая собой отношение с.к.о. V_1 и V_2 будет резко изменяться в начале и конце аномалии и без изменения в середине. В силу данного обстоя-

тельства, низкочастотная аномалия может фиксироваться как два коротких (высокочастотных) выброса. Для разрешения этой проблемы и для надежного определения такого рода аномалий, целесообразно использовать коэффициент Кохена – d_e. Коэффициент d_e равен разнице между двумя средними величинами $\bar{S}_1$ и $\bar{S}_2$ поделенной на среднеквадратическое отклонение выборки для данных средних величин

$$d_e = \frac{\bar{S}_1 - \bar{S}_2}{\sqrt{(V_1^2 + V_2^2)/2}}, \qquad (2)$$

где $\bar{S}_1$ и $\bar{S}_2$ – средние значения выборок в окнах обнаружения $\mathbf{W}_1$ и $\mathbf{W}_2$, соответственно, а V_1 и V_2 – среднеквадратическое отклонение указанных выборок.

Отношение d_e можно рассматривать как усредненный выигрыш, который будет характеризовать величину изменения выборки в окне обнаружения $\mathbf{W}_2$ по сравнению с общей выборкой. Главное состоит в том, что при обнаружении НЧ аномалии средний выигрыш будет стабильно больше чем 1,0. Следовательно, используя параметр d_e можно точно определить НЧ аномалию. Начальное детектирование аномалий подразумевает двухпороговый механизм обнаружения. По сути, речь идет о модифицированном критерии Вальда. Для первоначального обнаружения ВЧ-аномалий устанавливается два порога обнаружения: порог η_1, указывающий на однозначное наличие аномалии и порог η_2, указывающий на необходимость дальнейшей декомпозиции временного ряда, причем $\eta_1 > \eta_2$. Для начального обнаружения НЧ – аномалии устанавливается порог ее наличия – θ_1 и порог декомпозиции для дальнейшего поиска аномалий – θ_2, причем $\theta_1 > \theta_2$. Очевидно, что если порог наличия аномалии достигнут (η_1 или θ_1), то предполагается что она есть. Если же достигнут только порог декомпозиции (η_2 или θ_2), то полагается что аномалия может иметь место. Первоначально выполняется кратномасштабная вейвлет-декомпозиция выборки сигнального трафика для уровня 1. Далее совершается пакетная вейвлет-декомпозиция сигнала от узла [1,0] до 3-го уровня. После вейвлет-декомпозиции можно обнаруживать аномальные коэффициенты при различных значениях масштаба используя статистические алгоритмы обнаружения со скользящими окнами (1, 2). Когда обнаружено, что значение аномального выброса достигло порога декомпозиции на определенном уровне n-го масштаба разложения, то после этого сразу начинается этап обнаружения на уровне реконструкции сигнала.

Если аномалия и в этом случае имеет место, то фиксируется сигнал тревоги. Когда обнаружено, что аномалия достигла порога декомпозиции на определенном уровне n-го масштаба, то декомпозиция продолжается для обнаружения аномалий на n + 1 уровне и т.д. Декомпозиция будет закончена или в случае достижения порога тревоги или если уровень аномалии окажется ниже порога декомпозиции.

Алгоритм обработки выборочных данных сетевого трафика включает пять основных этапов обработки (рис. 1): получение выборки, пакетный вейвлет-анализ, первичное обнаружение аномалии трафика, реконструкция сигнала по вейвлет-коэффициентам и подтверждение наличия аномалии.

1. Создание исследуемого сигнала. Мы принимаем пакеты, собираемые маршрутизатором в единицу времени как сигнальный трафик с интервалом выборки T_0, сек. Если f(n) – значение n-й выборки, то тогда:

$$f(n) = \begin{cases} 0, & \text{при } n = 0. \\ (T_0 \cdot (n-1), T_0 \cdot n] & - \text{ количество} \\ \text{зарегистрированных IP-пакетов.} & \end{cases} \qquad (3)$$

2. Пакетный вейвлет-анализ сформированной выборки. Даже слабая высокочастотная аномалия обнаруживается в вершине (1,1) на первом уровне разложения пакетной вейвлет-декомпозиции.

Если процесс продолжить, то с увеличением числа уровней пакетного вейвлет-анализа количество коэффициентов уменьшится наполовину. Если длина обрабатываемого временного ряда N, то длина последовательности вейвлет-коэффициентов на уровне j будет равна $N/2j$. Кроме того, так как пакетное вейвлет-разложение базируется на степени 2 дискретного вейвлет-преобразования, то с увеличением числа уровней разложения количество вершин каждого уровня увеличивается в соответствии с 2^j. Следовательно, количество уровней декомпозиции с самого начала должно быть ограничено. Вообще предполагается применение адаптивной декомпозиции, в зависимости от ситуации, которая складывается при обнаружении.

3. Первоначальное обнаружение аномалий. Первоначальное обнаружение аномалий трафика осуществляется на коэффициентах пакетного вейвлет-разложения при каждом значении масштаба. При этом, с помощью статистических выражений Z и d_e проверяем имеются ли аномалии трафика при данном значении масштаба. Если $Z > з_1$ или $d_e > и_1$, что является признаком аномалии, то осуществляется переход к шагу 4. В случае, если отношение $Z \le \eta_1$, $Z > \eta_2$ или $d_e \le \theta_1$, $d_e > \theta_2$, что говорит о возможности наличия аномалии, осуществляется пакетная вейвлет-декомпозиция сигнала и производится операция обработки в соответствии с шагом 3

повторно. Если отношение $Z \le \eta_2$ и $d_e \le \theta_2$, то возможная аномалия не подтвердилась. Таким образом, можно утверждать, что на уровне декомпозиции проявляются свойства самоадаптации.

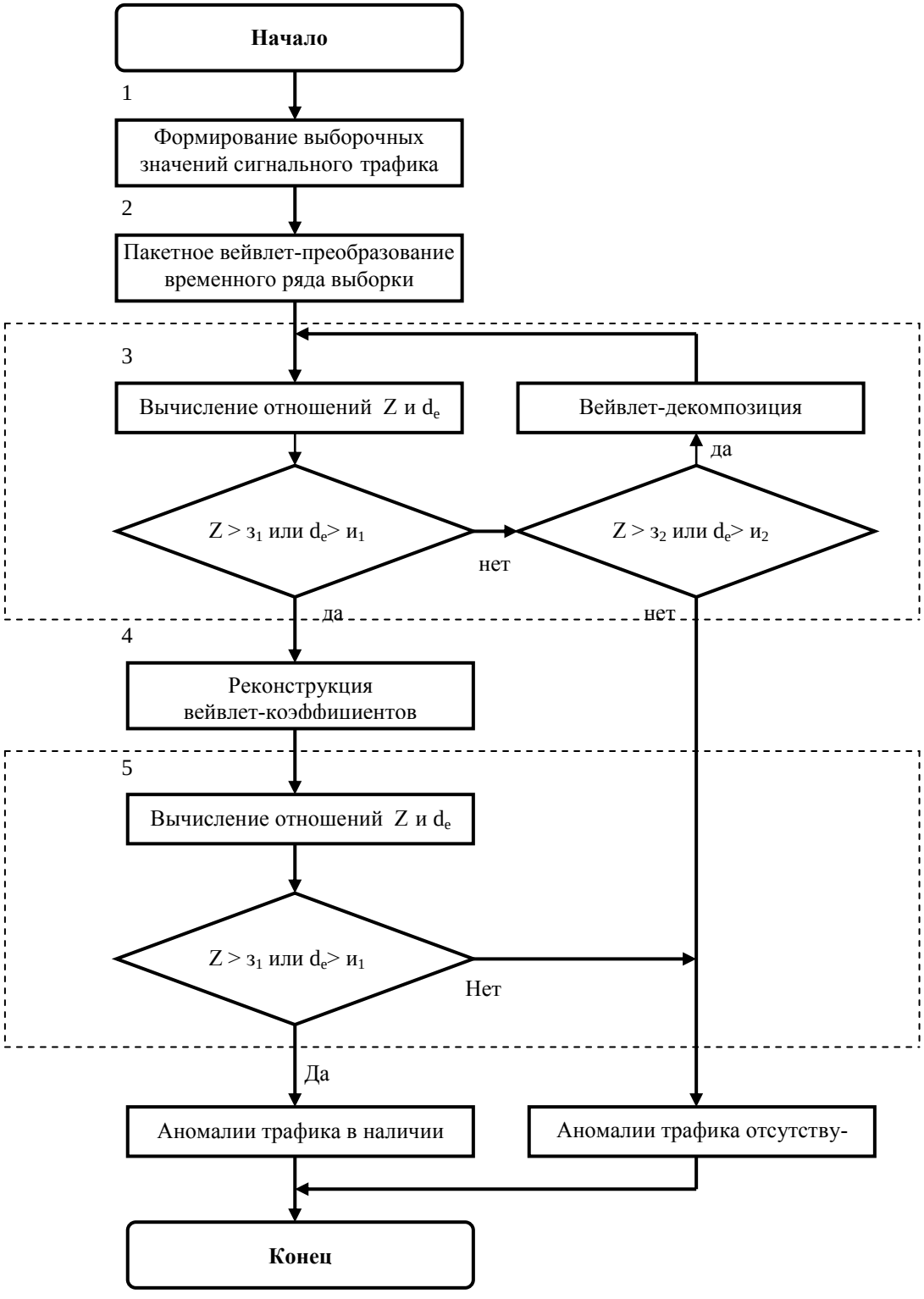

Рис. 1 Алгоритм обработки выборки сигнального трафика для обнаружения аномалий в сети

4. Реконструкция сигнала вейвлет-пакетом. Энергия аномальных сигналов, в основном, будет сконцентрирована на тех значениях масштаба, при которых, на шаге 3, были обнаружены аномалии. Следовательно, в дальнейшем, необходимо осуществить избирательную реконструкцию именно этих вейвлет-коэффициентов. В результате восстановления, полученный сигнал может отличаться от исходного сигнала. В частности, при этом происходит фильтрация вейвлет-коэффициентов и удаляются шумы. Для фильтрации коэффициентов детализации, выполняемой на четвертом этапе, лучше ис-

пользовать метод "мягкой" пороговой фильтрации. При этом коэффициенты, абсолютное значение которых меньше порогового, обнуляются, а остальные – "подтягиваются" к нулевому значению на величину порога:

$$y = \begin{cases} x + \theta, & \text{если } x < 0 \text{ и } |x| > \theta; \\ x - \theta, & \text{если } x > 0 \text{ и } |x| > \theta; \\ 0, & \text{если } |x| \le \theta, \end{cases} \quad (4)$$

где x – значение коэффициента до фильтрации; y – значение коэффициента после фильтрации; и – порог.

5. Подтверждение наличия аномалии в сети. Данные об аномалии, которые получаются на первоначальном этапе детектирования, могут оказаться ложными. Поэтому необходима повторная обработка реконструированного сигнала для уменьшения вероятности ошибки. Если в результате повторной обработки реконструированного сигнала имеет место превышения порога тревоги, то подтверждается наличие аномалии, в противном случае устанавливается наличие ошибки.

Выводы

1. Предложенный метод обнаружения основывается на вейвлет-представлении временного ряда сигнального трафика и статистическом алгоритме обнаружения с двумя порогами.

2. В данном методе впервые применен статистический критерий Кохена, позволяющий эффективно выявлять низкочастотные аномалии.

3. Разработанный метод может быть применен для класса задач обнаружения аномалий в компьютерных и телекоммуникационных сетях.

Список литературы

1. Lewis L. Extending trouble ticket systems to fault diagnosis / L. Lewis, G. Dreo // IEEE Network. – 1993. – Vol. 7. – P. 44-51.

2. Lewis L. A case based reasoning approach to the management of faults in communication networks / L. Lewis // Proc. IEEE INFOCOM. – 1993. – Vol. 3. – P. 1422-1429.

3. Katzela I. Schemes for fault identification in communication networks / I. Katzela, M. Schwarz // IEEE/ ACM Trans. Networking. – 1995. – Vol. 3. – P. 753-764.

4. Rouvellou I. Automatic alarm correlation for faultidentification / I. Rouvellou, G. Hart // Proc. IEEE INFOCOM. – 1995. – P. 553-561.

5. Feather F. Fault detection in an Ethernet network using anomaly signature matching / F. Feather, R. Maxion. – 1993. – Vol. 23. – P. 279-288.

6. Papavassiliou S. Implementing enhanced network maintenance for transaction access services / S. Papavassiliou, M. Pace, L. Ho // Tools and applications, 2000. – Vol. 1. – P. 279-288.

7. Thottan M. Anomaly detection in ip networks / M. Thottan, J. Chuanyi // IEEE TRANSACTIONS ON SIGNAL PROCESSING. – 2003. – Vol. 51, № 8. – P. 279-288.

8. Cheng C.-M. Use of spectral analysis in defense against dos attacks / C.-M. Cheng, H. Kung, K.-S. Tan // Proceedings of IEEE GLOBECOM 2002. – 2002. – P. 279-288.

9. Allen W. the self-similary of synthetic traffic for the evaluathion of intrusion detection system / W. Allen, G. Marin // Proceedings of the 2003 Symposium on Applications and the Internet (SAINT03). – 2003. – P. 279-288.

10. Lakhina A. Diagnosing network-wide traffic anomalies / A. Lakhina, M. Crovella, C. Diot // ACM SIGCOMM. – 2004. – P. 279-288.

11. Alarcon-Aquino V. Anomaly detection in communica-tion networks using wavelets / V. Alarcon-Aquino, A. Barria // IEEE Proc-Commun. – 2001. – Vol. 148, № 6. – P. 279 – 288.

Поступила в редколлегию 18.02.2009

Рецензент: д-р техн. наук, проф. В.Ф. Олейник, Национальная комиссия по вопросам регулирования связи Украины, Киев.

МЕТОД ВИЯВЛЕННЯ АНОМАЛІЙ ПОТОКІВ ДАНИХ В МЕРЕЖАХ

Г.О. Максименко

Пропонується метод для вирішення завдань виявлення аномалій трафіка в комп'ютерних та телекомунікаційних мережах, в основі якого лежить пакетне вейвлет-перетворення даних трафіка та статистичний алгоритм виявлення, що використовує критерій Кохена. Критерій Кохена дозволяє понизити рівень помилкових тревог при виявленні аномалій трафіка на низьких частотах. Даний алгоритм використовує два переміщуваних вікна для виявлення аномалій, два порогових значення та відрізняється від інших відомих низьким рівнем помилкових тривог, високою швидкістю обробки та адаптивністю.

Ключові слова: *виявлення, аномалія, трафік, вейвлет-перетворення.*

ANOMALY TRAFFIC DETECTION METHOD IN COMMUNICATION

G.A. Maksimenko

At this articles proposed method for detection anomaly traffic in communication and computer network. This method based on packet wavelet-decomposition signal traffic and detection statistical algorithm. This method have high speed calculation, lower level false alarm and adaptation capability.

Keywords: *finding, anomaly, traffic, wavelet-transformation.*