УДК 004.9

Д.К. Левоневский, Ю.А. Пичугин, Р.Р. Фаткиева

ОЦЕНКА СПЕКТРАЛЬНЫХ ХАРАКТЕРИСТИК ТРАФИКА В ЗАДАЧЕ ОБНАРУЖЕНИЯ КОМПЬЮТЕРНЫХ АТАК РАЗЛИЧНОГО ТИПА

Левоневский Д.К., Пичугин Ю.А., Фаткиева Р.Р. **Оценка спектральных характеристик трафика в задаче обнаружения компьютерных атак различного типа.**
Аннотация. В статье рассматриваются оценки чувствительности спектральных характеристик анализируемого входного трафика к компьютерной атаке при использовании альтернативы сингулярного спектрального анализа типа «гусеницы» для различных метрик и видов атак. Выявленное изменение спектра в момент начала атаки и при её продолжении может использоваться при разработке средств обнаружения вторжений.
Ключевые слова: информационная безопасность, Distributed Denial of Service, сетевой трафик, сингулярное спектральное разложение.

Levonevskiy D.K., Pichugin Y.A., Fatkieva R.R. **Estimation of traffic spectrum characteristics in the problem of various computer attacks detection.**
Abstract. The paper considers estimations of the incoming network traffic spectrum characteristics sensitivity to computer attacks. The spectrum is built by means of singular spectrum analysis and “the caterpillar” for various attacks and traffic functions. The discovered spectrum change at the moment of the beginning of an attack and during its running can be useful for intrusion detection systems development.
Keywords: information security, Distributed Denial of Service, DDoS, network traffic, singular spectrum analysis.

1. Введение. Задача обнаружения несанкционированного доступа к Web-серверам или компьютерных атак предполагает, в свою очередь, проведение различного рода исследований по определению косвенных признаков наличия или отсутствия атаки [1]. Одним из наиболее перспективных направлений таких исследований представляется спектральный анализ входных и выходных трафиков, которые в данном контексте рассматриваются как отрезки временных рядов [2]. При этом особое место занимают алгебраические методы, позволяющие относительно быстро выделять частотные составляющие, на которые ложится большая часть общей дисперсии. К таковым относятся хорошо известные методы, такие, как анализ сингулярного спектра (ASS – singular spectrum analysis) и метод «гусеницы», отличающиеся друг от друга способом вычисления матрицы автокорреляций R [3]. Целью настоящей работы является оценка чувствительности спектральных характеристик анализируемого входного трафика к компьютерной атаке при использовании метода

«гусеницы» для различных метрик и видов атак. При этом конечный результат работы предполагает также и выбор метрик инвариантно реагирующих на любой тип атаки.

2. Выбор оцениваемых спектральных характеристик трафика. Основной гипотезой настоящего исследования является предположение, что спектры, т.е. собственные значения автокорреляционной матрицы, должны реагировать на появление компьютерной атаки. При этом следует учесть, что размерность автокорреляционной матрицы может быть достаточно большой. В методе ASS она равна длине анализируемого отрезка временного ряда, а в методе «гусеницы» – половине длины этого отрезка. В связи с этим целесообразно рассмотреть какие-либо общие спектральные характеристики, позволяющие характеризовать изменение спектра в целом, а с другой стороны, достаточно чувствительные к наличию атаки.

В качестве анализируемых спектральных характеристик возьмем количество статистически значимых главных компонент (КСЗК) [4], которое в формулах обозначим через p, и отношение первого (наибольшего) собственного значения к общей сумме спектральных чисел – $\lambda_1 / \sum \lambda_i = \lambda_1 / \operatorname{tr} R$ (tr – след матрицы). В силу нормирования $\sum \lambda_i = \operatorname{tr} R = M$, где M – размерность (длина «гусеницы»). Поэтому можно использовать (и писать) λ_1 / M. Относительно определения величины КСЗК необходимо сделать некоторые пояснения. При анализе главных компонент проверяется гипотеза:

$$H_0 : \lambda_1 > \lambda_2 > \ldots > \lambda_p > \lambda_{p+1} = \lambda_{p+2} = \lambda_{p+3} = \ldots = \lambda M ,$$

где «хвост» равных собственных значений (т.е. фактическое различие считается случайным), со статистической точки зрения, относится к неинформативной «шумовой» части спектра. Известно, что при верной гипотезе H_0 величина [2]:

$$\gamma_r = (k-1)\left((M-p)\ln\left(\frac{1}{M-p} \sum_{i=p+1}^{M} \lambda_i \right) - \sum_{i=p+1}^{M} \ln\lambda_i \right),$$

имеет распределение χ^2_η с числом степеней свободы $\eta = r(r+1)/2 - 1$, где $r = M - p$ количество элементов, которые относятся к «шуму» (длина «хвоста»). Это означает, что в случае, когда тестовое значение γ_r превышает пороговое значение $\chi^2_\eta(\alpha)$ для некоторого наперед

заданного уровня значимости α, например 0,05 или 0,01, мы должны отбросить гипотезу H_0, т.е. увеличить значение p (укоротить «хвост»). Однако, как показала практика, использование этого теста нередко приводит к весьма завышенному значению p. В связи с этим последнее время стал более популярен тест, известный, как правило «сломанной трости» (Broken stick model [5]). Согласно этому правилу, p равняется максимальному значению, для которого выполняется неравенство:

$$\lambda_p / trR > (1/p + 1/(p+1) + ... + 1/M)/M ,$$

а учитывая замечание относительно нормирования ($\sum \lambda_i = \mathrm{tr}\, R = M$),

$$\lambda_p > 1/p + 1/(p+1) + ... + 1/M .$$

3. Анализ результатов численного эксперимента. В качестве исходных временных рядов в численных экспериментах взяты значения метрик — специальных функций системного трафика. Список из десяти использованных метрик приводится в таблице.

Таблица 1. **Метрики, идентифицирующие атаку**

Метрики	Типы атак	Комментарии
R_{ip} – отношение объёмов входящего и исходящего трафика; R_{nip} – отношение числа или количества входящих и исходящих пакетов	Все	Характеризует способность сервера отвечать на запросы
D_{ack} – разность исходящих и входящих флагов ACK	HTTP flood, SYN flood	Характеризует способность сервера отвечать на запросы
R_{udp} – доля UDP-пакетов в IP-трафике	SYN flood, UDP flood, HTTP flood	Характеризует степень загрузки сети однонаправленным UDP-трафиком
R_{syn} – отношение числа SYN-флагов к числу входящих пакетов; R_{psh} – отношение числа PSH-флагов к числу входящих пакетов; R_{sp} – отношение числа SYN-флагов к числу PSH-флагов	SYN flood	Характеризуют эффективность передачи данных (Rpsh показывает долю полезной нагрузки, Rsyn - долю накладных расходов, Rsp является интегрированным показателем)
R_{tcp} – эффективность TCP-протокола	SYN flood, TCP flood	Показывает степень использования TCP для передачи данных прикладным программам
L_{avg} – средняя длина пакета; L_{tcp} – средняя длина TCP-пакета	SYN flood, ICMP flood	Определяет средний объём передаваемых данных

Как было отмечено выше, задачей настоящего исследования является выбор метрик, при использовании которых выбранные спектральные характеристики инвариантно реагируют на появление нелегального трафика (атаки).

В первую очередь следует отметить, что на атаку типа SYN flood все метрики без исключения реагируют увеличением анализируемой спектральной характеристики (рис. 1–3).

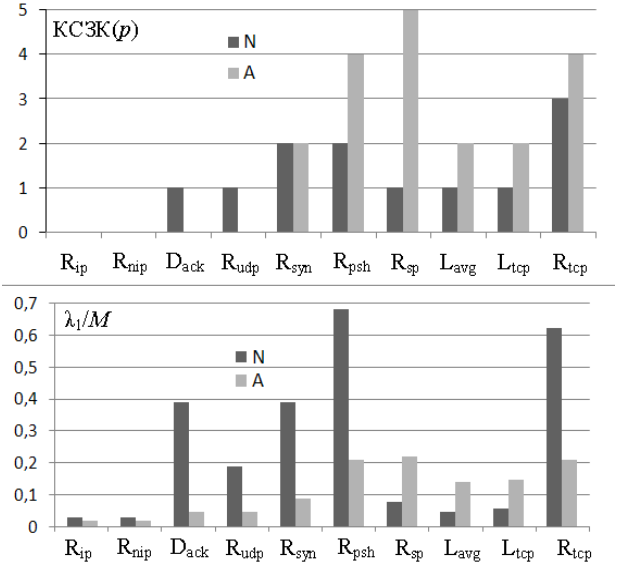

Рис. 1. Реакция спектральных характеристик входного трафика на атаку типа HTTP flood (N – до атаки; A – во время атаки).

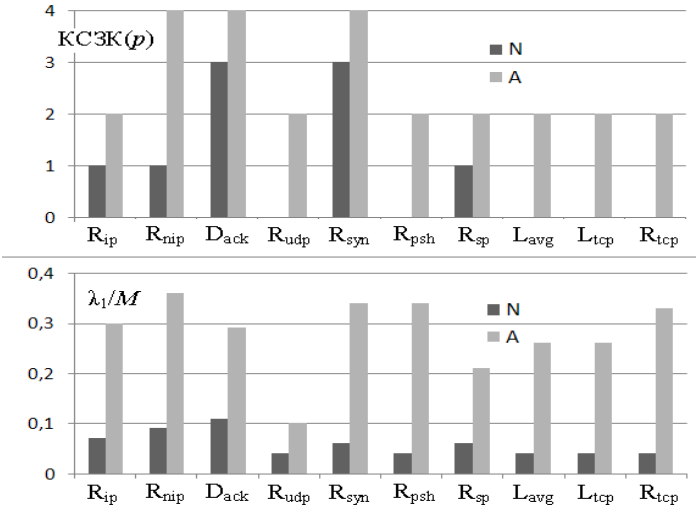

Рис. 2. Реакция спектральных характеристик входного трафика на атаку типа SYN flood (N – до атаки; A – во время атаки).

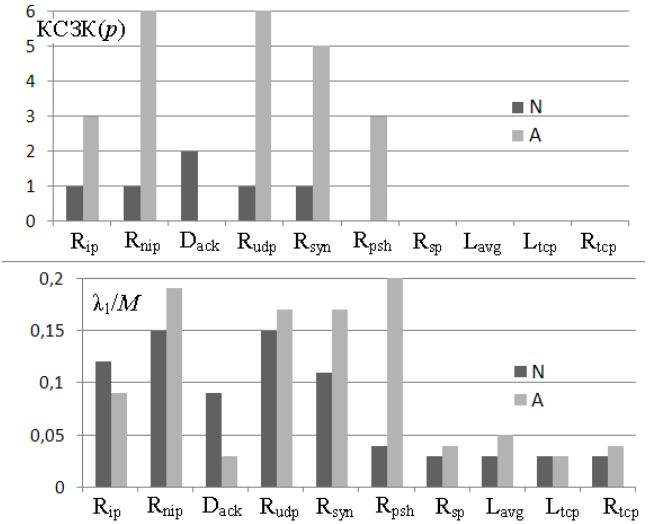

Рис. 3. Реакция спектральных характеристик входного трафика на атаку типа UDP flood (N – до атаки; A – во время атаки).

4. Заключение. На основании проведенного исследования можно сделать вывод, что выбранные спектральные характеристики, несомненно, могут быть использованы в качестве индикаторов (косвенных признаков) для обнаружения компьютерной атаки. Результаты численных экспериментов позволяют выделить метрики, реагирующие на все типы компьютерных атак.

К общему выводу следует добавить целесообразность в рамках данной проблемы дальнейших исследований по использованию алгебраических методов спектрального анализа. Здесь могут быть поставлены такие вопросы, как влияние загруженности трафика на чувствительность выбранных критериев, выбор оптимальной длины анализируемых отрезков трафика, а также использование иных спектральных характеристик и т.п.

Литература

1. *Котенко И.В., Юсупов Р.М.* Текущее состояние и тенденции развития в области построения безопасных компьютерных систем // Часть 5-й Российской мультиконференции по проблемам управления (МКПУ-2012) - конференция "Информационные технологии в управлении" (ИТУ-2012). 09–11 октября 2012 г. Материалы конференции. СПб, 2012. С. 671–675.
2. *Фаткиева Р.Р., Левоневский Д.К.* Детектирование компьютерных атак методом сингулярного спектрального разложения // Труды СПИИРАН. 2013. Вып. 25. С. 135–147.
3. *Данилов Д.Л., Жиглявский А.А.* (ред.) Главные компоненты временных рядов: метод «Гусеница». СПбГУ, 1997. 308 с.
4. *Левоневский Д.К., Пичугин Ю.А., Фаткиева Р.Р.* Исследование компьютерных атак методом сингулярного спектрального разложения сетевого трафика // Труды СПИИРАН. 2013. Вып. 26. С. 101–114.
5. *Bartkowiak A.* How to reveal the dimensionality of the data? // Applied Stochastic Models and Data Analysis. 1991. С. 55–64.

Левоневский Дмитрий Константинович — бакалавр информационных систем, младший научный сотрудник лаборатории информационно-вычислительных систем СПИИРАН. Область научных интересов: исследование DDoS-атак, статистический анализ и моделирование трафика локальных сетей. Число научных публикаций — 4. DLewonewski.8781@gmail.com; СПИИРАН, 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-4369, факс +7(812)328-4450.

Levonevskiy Dmitriy Konstantinovich — bachelor on information systems, researcher, Laboratory of Computer and Information Systems, SPIIRAS. Research interests: research of DDoS attacks, statistical analysis and modeling of the network traffic. The number of publications — 4. DLewonewski.8781@gmail.com; SPIIRAS, 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-4369, fax +7(812)328-4450.

Пичугин Юрий Александрович — д.ф.-м.н.; профессор кафедры прикладной математики Российского педагогического университета (РГПУ) им. А.И. Герцена.

Область научных интересов: разработка и использование статистических и динамических моделей для анализа и прогноза нестационарных многомерных временных рядов. Число научных публикаций — 81; РГПУ им. А.И. Герцена, ул. Казанская 6, 191186, Санкт-Петербург, РФ; р.т. +7(812) 314-48-85.

Pichugin Yury Alexandrovich — Ph.D., Dc.Sci.; Professor of Department of Applied Mathematics, Herzen State Pedagogical University. Research interests: designing and use of statistical and dynamical models for multidimensional non-stationary time-series analysis and forecast. The number of publications — 81; yury-pichugin@mail.ru; Herzen State Pedagogical University, 6 Kazanskaya st. 191186, St. Petersburg, Russia; office phone +7(812) 314-48-85.

Фаткиева Роза Равильевна — к.т.н.; старший научный сотрудник лаборатории информационно-вычислительных систем СПИИРАН. Область научных интересов: моделирование информационных систем. Число научных публикаций — 30. rikki2@yandex.ru; СПИИРАН, 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-4369, факс +7(812)328-4450.

Fatkieva Rosa Ravilievna — Ph.D., senior researcher, Laboratory of Computer and Information Systems, SPIIRAS. Research interests: modeling of information systems. The number of publications — 30. rikki2@yandex.ru; SPIIRAS, 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-4369, fax +7(812)328-4450.

Рекомендовано лабораторией информационно-вычислительных систем СПИИРАН, заведующий лабораторией Воробьев В.И, д.т.н., проф.
Статья поступила в редакцию 16.08.2013.

.

РЕФЕРАТ

Левоневский Д.К., Пичугин Ю.А., Фаткиева Р.Р. **Оценка спектральных характеристик трафика в задаче обнаружения компьютерных атак различного типа.**

Задача обнаружения несанкционированного доступа к информационным системам предполагает проведение исследований по определению косвенных признаков наличия или отсутствия атаки. Одним из наиболее перспективных направлений исследования представляется спектральный анализ сетевого трафика. При этом особое место занимают алгебраические методы, позволяющие относительно быстро выделять частотные составляющие, на которые ложится большая часть общей дисперсии. К ним относятся хорошо известные методы, такие, как анализ сингулярного спектра и метод «гусеницы», отличающиеся друг от друга способом вычисления матрицы автокорреляций. Оба метода позволяют характеризовать изменение спектра в целом, и достаточно чувствительны к наличию атаки.

Рассмотренный в статье анализ спектральных характеристик входного трафика к компьютерной атаке при использовании метода «гусеницы» для различных метрик и видов атак показал, что для многих измеряемых метрик при появлении атаки количество статистически значимых компонент (КСЗК) спектра возрастает на несколько единиц и возрастает относительная величина первого (наибольшего) собственного значения (λ_1 / M , M – размерность).

Исходя из результатов численных экспериментов, определены метрики, дающие реакцию на любой вид компьютерной атаки (для КСЗК — R_{psh}, для λ_1 / M — L_{avg} и L_{tcp}). Все анализируемые метрики разбиты на группы в зависимости реакции на тип атаки. При этом установлено, что на атаку типа SYN flood реагируют все метрики без исключения.

На основании проведенного исследования можно сделать вывод, что спектральные характеристики входного трафика, несомненно, могут быть использованы в качестве индикаторов (косвенных признаков) для обнаружения компьютерной атаки.

SUMMARY

Levonevskiy D.K., Pichugin Y.A., Fatkieva R.R. **Estimation of traffic spectrum characteristics in the problem of various computer attacks detection.**

The problem of detection of illegal access to information systems requires to perform supplementary research in order to reveal indirect features of a DDoS attack. One of the most prospective research directions seems to be the traffic spectrum analysis. Furthermore algebraic techniques, that enable figuring out frequency constituents containing the most part of the total variance quickly enough, have a significant place. Well known techniques of singular spectrum analysis and “the caterpillar”, differing by the method of autocorrelation matrix computing, belong to this kind of methods. Both methods allow to describe the change of the spectrum as a whole, and are quite sensitive to the presence of an attack.

Considered in this paper analysis of the spectral characteristics of the input traffic, performed to figure out their sensitivity to a computer attack by means of the "caterpillar" method for various metrics and types of attacks, showed that the number of the statistically significant major constituents for a lot of the measured metrics increases by several units after the beginning of an attack. The first (the greatest) relative eigenvalue (λ_1 / M, M is the dimension) increases also.

Considering the results of the numerical experiments, we determined the metrics reacting upon all concerned kinds of DDoS attacks (R_{psh} for the number of the statistically significant major constituents, L_{avg} and L_{tcp} for λ_1 / M). All analyzed metrics are divided into groups depending on their reaction upon attacks. At the same time it is revealed that all metrics, without exception, react upon the SYN flood.

On the basis of the performed research we can draw a conclusion that the incoming traffic spectrum characteristics can surely be used as detectors (indirect indicators) for computer attacks.