УДК: 004.7.056.53

ПРИМЕНЕНИЕ СТАТИСТИЧЕСКИХ МЕТОДОВ ОБНАРУЖЕНИЯ DoS АТАК В ЛОКАЛЬНОЙ СЕТИ

Н. А. Семенов, А. Ю. Телков

Воронежский государственный университет

Поступила в редакцию 04.05.2012 г.

Аннотация. В настоящей статье исследуются возможности применения статистических алгоритмов для обнаружения DoS-атак на локальную сеть с использованием функционала стандартной ЭВМ с двумя сетевыми интерфейсами, работающей под управлением ОС Linux. Предложен и практически реализован алгоритм, позволяющий определить оптимальное значение величины локального профиля, и не требующий допущения о нормальности сетевого трафика.

Ключевые слова: DoS-атака, скользящая средняя, UDP-наводнение, обнаружение сетевых аномалий.

Annotation. In this paper, we continue investigating capabilities of applying statistical algorithms for DoS attacks detection in local networks based on functionality of standart host machine with two network interfaces and Linux OS. Algorithm for determining optimal size value of local profile, not dependent on network traffic normality assumption, was proposed and practically implemented.

Key words: DoS attack, moving average, UDP-flood, network anomaly detection.

ВВЕДЕНИЕ

Бурное развитие сетевых технологий и сети Интернет приводит к росту числа нарушений, связанных с информационной безопасностью. Согласно годовому международному отчету о безопасности сетевой инфраструктуры (Network Infrastructure Security Report) [1], представленному в феврале 2011 года, 2010 год ознаменовался резким ростом как числа DDoS-атак, так и их интенсивностью. Наибольшая зафиксированная интенсивность DDoS-атаки в 2010 году составила 100 гигабайт в секунду, превысив показатель 2009 года более чем на 100%, а показатель 2005 года – более чем на 1000%. DDoS-атаки стали самым распространенным типом компьютерных угроз, причем такие типы DDoS-атак как UDP и ICMP наводнение являются наиболее крупными по своей интенсивности.

В данной работе продолжается исследование различных статистических подходов по обнаружению нарушений безопасности в компьютерных сетях. По отмеченным выше причинам, особенный интерес для нас представляло обнаружение DDoS-атак типа UDP-наводнение, как одних из самых распространенных и разрушительных для компьютерных сетей. [1] В качестве платформы для сбора статистики и анализа сетевого трафика использовался стандартный Linux-шлюз – компьютер с двумя сетевыми интерфейсами, выполняющий функции пакетной фильтрации и маршрутизации в сетях многих предприятий и организаций.

Статистический подход, представленный нами ранее, позволял обнаруживать аномальные изменения в трафике, характерные для DDoS-атак, но имел ряд недостатков. Во-первых, не было предложено алгоритма выбора оптимального значения для величины локального профиля, которое используется для расчета выборочного среднего; во-вторых, делалось допущение о нормальном распределении сетевого трафика – условие, которое в современных, динамично изменяющихся сетевых условиях и топологиях редко выполняется. В данной работе предлагается использовать метод контрольных карт EWMA для статистического анализа сетевого трафика – данный подход, согласно [2], [3], [4], не требует допущений о нормальности сетевого трафика. Кроме того, была представлена методика нахождения оптимального размера локального профиля, используя методологию EWMA.

В части 2 данной работы представлено теоретическое описание используемого подхода,

описана методика нахождения оптимального коэффициента EWMA, а также рассмотрен механизм злонамеренного обучения системы обнаружения вторжений. В части 3 приводятся описание и результаты проводимого эксперимента по обнаружению DDoS-атак типа UDP-наводнение в реальной сетевой топологии с использованием Linux-шлюза. В части 4 сделаны выводы о проделанной работе, и рассмотрены возможные будущие задачи.

2. ИСПОЛЬЗОВАНИЕ КОНТРОЛЬНЫХ КАРТ EWMA ДЛЯ ОБНАРУЖЕНИЯ ВТОРЖЕНИЙ В СЕТЬ

Контрольные карты EWMA (Exponentially Weighted Moving Average – скользящая средняя с экспоненциальным распределением весов) были впервые введены Робертсом в 1959 году [4], и с тех пор нашли применение в управлении бизнес-процессами, производством, для статистического контроля качества в различных технологических процессах.[5] Свойства алгоритма EWMA были аналитически оценены в работе Лукаса и Саккьюзи.[6]

EWMA-статистика используется для контроля процессов, в ходе которых происходит усреднение значений выборки некоторой случайной величины таким образом, что вес каждого из значений выборки убывает с течением времени, и со временем они перестают учитываться. [7] Для техники контрольных карт EWMA решение о том, находится ли случайный процесс в обычном или аномальном состоянии, зависит от статистики EWMA, которая представляет собой скользящую среднюю с экспоненциальным распределением весов для всех значений случайного процесса, включая вновь получаемые.

Отметим, что применение алгоритма EWMA к задаче обнаружения вторжений в компьютерные сети рассматривалось исследователями в [8]. Были получены результаты, подтверждающие пригодность применения процедуры контроля EWMA для данной задачи. Авторами использовался лишь один порог, однократное превышение которого расценивалось как аномалия. Порог не был зафиксирован, и система обнаружения, таким образом, была подвержена фундаментальному недостатку – злонамеренному обучению злоумышленником.

В данной работе предлагается верхний контрольный порог зафиксировать для избежания злонамеренного обучения, а для сохранения системой возможности адаптивного реагирования на суточные, недельные флуктуации трафика, вводится дополнительный порог, с которым сравниваются определенное количество последних значений трафика. Выбор их числа зависит от оптимальных значений параметра EWMA и размера локального профиля, процедура нахождения которых описана в этой работе.

Выбирая весовой коэффициент λ, можно менять чувствительность процедуры контроля EWMA в сторону обнаружения как больших, так и малых, постепенных отклонений значений процесса.

Согласно [4], формула для расчета статистики EWMA имеет следующий вид:

$$z_t = \lambda X_t + (1-\lambda) z_{t-1}, \qquad (1)$$

где λ – весовая константа, принимающая значения от 0 до 1 ($0 < \lambda \le 1$) и определяющая величину памяти EWMA; t – время наблюдения ($t \in 0,1,2..n$); n – число наблюдений; z_t – значение статистики EWMA в момент времени t; X_t – значение наблюдаемой величины в момент времени t.

Процедуру контроля с помощью контрольных карт EWMA можно применить для обнаружения аномалий в сетевом трафике компьютерной сети предприятия, которые могут свидетельствовать о DoS-атаке на данную сеть.

Если рассматривать в качестве случайной величины число величину входящего трафика в данном сегменте сети за равные промежутки времени, то каждое новое значение этой величины можно сравнивать с пороговым значением. Величина порогового значения меняется адаптивно, на основании оценки среднего значения входящего трафика, в соответствии с суточными и недельными колебаниями уровня сетевой активности. Если величина порога превышена, в момент времени n будет объявлена «тревога»:

$$X_n \ge (\beta + 1)\mu_{n-1}, \qquad (2)$$

где X_n – это значение величины входящего трафика (байт в секунду) в момент времени n, μ_{n-1} – значение среднего для всех значений процесса, предшествующих n; $0 < \beta < 1$ – пороговый коэффициент, который показывает, на сколько процентов текущее значение процесса должно превысить величину среднего, чтобы ситуация в сети считалась аномальной.

Используя статистику EWMA для предыдущих $n-1$ измерений величины входящего трафика, можем рассчитать среднее в момент времени n:[9]

$$\mu_n = (1-\lambda)X_n + \lambda\mu_{n-1}, \quad (3)$$

где λ – коэффициент EWMA.

Однократное превышение порогового значения не может служить однозначным доказательством наличия аномалии. Для того чтобы снизить количество ошибок первого рода (ложных срабатываний), алгоритм лучше настроить таким образом, чтобы аномалия фиксировалась после определенного числа k последовательных превышений порога. Сигнал тревоги в этом случае будет подаваться, если:

$$\sum_{v=n-k+1}^{n} 1_{(X_v \geq (\beta+1)\mu_{v-1})} \geq k,\ k > 0 \quad (4)$$

Требуется найти способ определения оптимальных значений коэффициента EWMA, порогового коэффициента и n – величину локального профиля (или размер окна скользящей средней) – число значений последних n значений выборки, для которых считается скользящая средняя.

Параметр λ, который называется коэффициентом EWMA, определяет скорость, с которой более старые значения процесса исключаются из процесса расчета статистики EWMA по формуле (1), принимает значения от о до 1 ($0 < \lambda \leq 1$). Значение $\lambda = 1$ означает, что лишь последние полученные результаты измерений оказывают влияние на Z_t. В этом предельном случае, контрольные карты EWMA совпадают с контрольным картами Шухарта [10]. Таким образом, выбор значения λ влияет на величину статистического веса для более старых или более новых результатов измерений, и, таким образом, является компромиссом между требуемым уровнем чувствительности обнаружения аномалий и уровнем ошибок первого рода (ложные срабатывания).

В литературе приводятся лишь общие рекомендации по выбору этого параметра. Так, например, в [11] для контроля технологических процессов авторы предлагают использовать $\lambda = 0.3$. В [12] предлагается диапазон значений λ между 0.2 и 0.3. В [13] отмечается, что для приложений по наблюдению за качеством обычно используются значения λ в диапазоне от 0.05 до 0.25, и несколько большие значения для прогноза и контроля технологических процессов.

В настоящей работе, как уже отмечалось ранее, интерес представляет нахождение точного значения величины локального профиля n, т.е. числа крайних полученных измерений, для которых рассчитывается скользящая средняя. Согласно [4], экспоненциальное сглаживание может быть аппроксимировано скользящей средней с величиной интервала усреднения (далее – размер окна) n. При известном значении λ, размер окна может быть найден как:

$$n = (2/\lambda) - 1. \quad (5)$$

Рассмотрим процесс нахождения оптимального значения λ для случая обнаружения аномалий сетевого трафика.

Согласно базовому уравнению экспоненциального сглаживания[12], значение μ_0 в любой момент времени t может быть найдено согласно приведенному ниже выражению:

$$S_t = (1-\lambda)S_{t-1} + \lambda X_{t-1}, \quad (6)$$

Таблица 1

К иллюстрации процесса нахождения оптимального значения λ

Время	Значения выборки – X_t	S_t – экспоненциально сглаженные значения	Ошибка $E_t = (X_t - S_t)$	Квадратичная Ошибка $ES_t = (X_t - S_t)^2$
1	X_1			
2	X_2	X_1	E_2	ES_2
3	X_3	S_3	E_3	ES_3
...	...	...	...	...
N	X_N	S_N	E_N	ES_N
				TES

где $0 < \lambda < 1$, а $t > 3$, S_t – экспоненциально сглаженное значение величины X в момент времени t. Значение λ меняется с некоторым шагом, например в одну сотую.

Оптимальным значением λ является такое, которое соответствует минимальному среднему квадратичных ошибок (далее – ASE).

TES – это сумма квадратичных ошибок для N значений выборки и $\lambda = 0.01$.Среднее квадратичных ошибок (ASE) найдем как:

$$ASE = TSE/(n-1),\ \lambda = 0.01. \quad (7)$$

Итеративная процедура вычисления λ повторяется для остальных значений с выбранным шагом. Из полученных в результате данного алгоритма значений ASE выбирается минимальное, и, таким образом, соответствующее ему оптимальное искомое значение λ.

Согласно [14], дисперсия статистики EWMA равна:

$$\sigma_z = (\lambda/(2-\lambda))\sigma^2, \quad (8)$$

где σ – величина стандартного отклонения.

Центральная линия для контрольных карт EWMA определяется как среднее Z_0 для всех значений выборки. Верхний и нижний контрольный предел вычисляются по формулам: [14]

$$UL = Z_0 - m\sigma_z \quad (9)$$

$$LL = Z_0 - m\sigma_z \quad (10)$$

В данном исследовании практическую роль играет лишь верхний контрольный предел *UL*, так как низкий уровень входящего в сетевой сегмент трафика не является аномалией и не представляет угрозы для функционирования сети.

Величина m в формулах (9) и (10) может быть принята равной трем (согласно правилу «трех сигма»), или выбрана с использованием таблицы соответствия [5].

Используя выражение (5), можем рассчитать размер окна n (число значений, использующихся при расчете скользящей средней). В соответствии с этой формулой, скользящая средняя будет рассчитываться для n крайних значений выборки за время t.

$$\mu_t^n = (X_t + X_{t-1} + X_{t-2} + ... + X_{t-(n-1)})/n. \quad (11)$$

Логику работы алгоритма настраиваем таким образом, что значение скользящей средней обновляется каждый n-ый временной интервал, тогда t=n, и можно переписать (13):

$$\mu_n^n = (X_n + X_{n-1} + X_{n-2} + ... + X_1)/n. \quad (12)$$

Сигнал о наличии аномалии будет подан, если пороговое будет превышено k раз. Принцип будет правильно работать, если:

$$k \le n. \quad (13)$$

В случае если последнее значение величины входящего трафика в единицу времени превысит верхний контрольный предел *UL*, то сигнал о наличии аномалии будет подан сразу.

3. ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Приведенный выше алгоритм был практически реализован с использованием функционала Linux-шлюза – компьютера с установленной ОС Fedora Core 14 и двумя сетевыми интерфейсами, который выполнял функции маршрутизации и пакетной фильтрации в приведенной топологии. На нем был установлен коллектор сетевого трафика IPCAD, [15], выполняющий захват пакетов на сетевых интерфейсах Linux-шлюза локальной сети. Используемая тестовая топология изображена на рисунке 1. Граничный маршрутизатор выступал в роли шлюза в глобальную сеть, Linux-шлюз разбивал локальную подсеть на две подсети – 192.168.10.0/24 (в которой находился атакующий хост, с которого организовывались DoS атаки), и 10.0.0.0/8 (в ней располагались атакуемые компьютеры, в том числе Хост 3, выполняющий роль файлового сервера локальной сети). (рис. 1) Для написания скрипта, выполняющего сбор и статистический анализ сетевого трафика, использовался язык Perl, широко использующийся для администрирования компьютерных сетей, и позволяющий решать широкий спектр задач.

Согласно обширному исследованию, проведенного в [16], существует два основных подхода к обучению систем обнаружения вторжений – в одном из случаев для начальной тренировки системы используется лишь заведомо свободный от аномальной активности

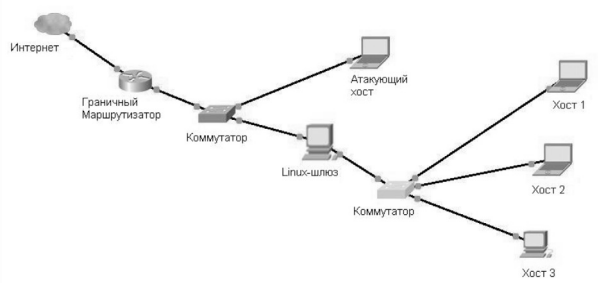

Рис. 1. Экспериментальная сетевая топология

трафик, в другом – под тщательным контролем, в процесс обучения включают различные типы сетевых атак. Мы считаем, что в случае задачи обнаружения DDoS-атак типа UDP-наводнение имеет смысл придерживаться первого подхода.

Рассмотрим на примере процесс нахождения оптимальных значений коэффициента EWMA λ и размера окна n, которые необходимы для вычисления верхнего и нижнего контрольных пределов, а также скользящей средней μ_n^n. Данные величины вычислялись для двенадцати значений входящего сетевого трафика (байты в секунду) усредненных за равные 3х-часовые промежутки времени, свободные от аномальной активности.

Для каждого значения λ от 0.01 до 1 с шагом в 0.01 вычисляем среднее квадратичных ошибок ASE, и определяем, какому значению λ соответствует минимальное ASE. Для значений сетевого трафика, полученных на стадии обучения системы и усредненных за 3 х-часовые промежутки времени значений входящего трафика были получены следующие результаты: минимальная сумма среднеквадратических ошибок ASE = 5364885105.91585, что соответствует значению коэффициента EWMA $\lambda = 0.34$. Данное полученное значение согласуется с рекомендованным диапазоном значений коэффициента EWMA для прогноза и контроля технологических процессов [11,12,13]. Следовательно, согласно формуле (5), оптимальный размер локального профиля (или размер окна скользящей средней) составляет 5 временных промежутков, n=5.

После вычисления значений λ, n, *UL*, констант m и β в рамках обучающей части скрипта, начинается его рабочий режим, в ходе которого в реальном времени происходит статистический анализ величины поступающего трафика. Как уже было отмечено выше, сигнал о наличии аномалии будет подан в двух случаях: незамедлительно – в случае, если последнее значение трафика превысило верхний контрольный предел *UL*; и, если последние n значений превысили величину адаптивного порога – скользящую среднюю $(\beta+1)\mu_{n-1}$ k раз, $k \le n$. Обе ситуации неоднократно имели место в рабочем режиме.

DoS-атаки типа UDP-наводнение, различной интенсивности, организовывались с атакующего хоста. Цель атакующего в случае UDP-наводнения – отправка в атакуемую сеть большого числа пакетов, что приводит к связыванию сетевых ресурсов данного сегмента сети – снижению пропускной способности канала и увеличению нагрузки на маршрутизаторы.

Аномальный трафик, вызванный организованными DoS-атаками, при прохождении через интерфейсы Linux-шлюза, обрабатывался разработанным скриптом в реальном времени. Сообщение об аномальной ситуации выдавалось на терминал Linux-шлюза через 5 временных интервалов после начала атак, $k \le n$, а в случае атак большой интенсивности – сразу после превышения фиксированного верхнего контрольного порога *UL*, что позволяет судить о пригодности данного подхода, основанного на использовании контрольных карт EWMA, для обнаружения аномальных изменений, вызываемых наличием DoS и DDoS атак в сетевом трафике.

ЗАКЛЮЧЕНИЕ

В данной работе продолжилось изучение методов статистического обнаружения аномалий сетевого трафика и возможности применения Linux-шлюза в качестве возможной платформы для системы обнаружения вторжений. Было рассмотрено практическое применение контрольных карт EWMA для задачи обнаружения аномалий в сети, вызванных проводимыми DoS и DDoS- атаками, на основе функционала стандартного Linux-шлюза. В результате применения данного подхода, во-первых, удалось отказаться от требования нормальности сетевого трафика (которое в современных масштабируемых компьютерных сетях является мало выполнимым). Во-вторых, в данной работе, значение величины локального профиля n (или размер окна), выбиралась оптимально, что сократило время обнаружения аномалий.

Однако, в ходе варьирования интенсивностей проводимых DoS-атак было отмечено, что существует тенденция ухудшения производительности алгоритма для атак малой интенсивности – было отмечено увеличение числа ложных тревог, и снижение вероятности обнаружения. Поскольку данный тип атак небольшой интенсивности, согласно [9], представляет наибольшую угрозу для компьютерных сетей, данный вопрос (зависимость производительности статистического алгоритма от характерис-

тик проводимых DoS-атак) требует дальнейшего изучения.

СПИСОК ЛИТЕРАТУРЫ

1. *Dobbins R., Morales C.* Worldwide Infrastructure Security Report // Worldwide Infrastructure Security Report, February 2011. URL: http://www.arbornetworks.com/report (дата обращения: 15.02.1012).

2. *Montgomery D. C.* Introduction to Statistical Quality Control / D.C. Montgomery//John Wiley and Sons, New York, NY, USA. – 1997.

3. *Bower K. M.* Using exponentially weighted moving average (EWMA) charts //Asia Pacific Process Engineer, October 2000. Систем. требования: Adobe Acrobat Reader. URL: http://www.mintab.com/uploadedFiles/Shared_Resources/Documents/Articles/ewma_control_charts.pdf (дата обращения: 15.02.1012).

4. *Roberts S. W.* Control charts based on geometric moving averages/ S.W. Roberts // Technometrics. – 1959. – Vol. 1. – P. 239–250.

5. *Neubauer A.S.* The EWMA Control Chart: Properties and Comparison with other Quality-Control Procedures by Computer Simulation/A.S.Neubauer// Clinical Chemistry. – 1997. – Vol. 43. – P. 594–601.

6. *Saccucci J. M., Lucas M. S.* Exponentially weighted moving average control schemes: Properties and enhancements/ J.M. Lucas ,M.S. Saccucci// Technometrics. – 1990. – Vol. 32. – P. 1–29.

7. Single Exponential Smoothing // NIST-SEMATECH e-Handbook of Statistical Methods, 2010. URL: http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc431.htm (дата обращения: 15.02.1012).

8. *Ye N., Borror C., Zhang Y.* EWMA techniques for computer intrusion detection through anomalous changes in event intensity /N. Ye, C. Borror, Y. Zhang //Quality and Reliability Engineering International. – 2002; –Vol.18. – P. 443–451.

9. *Vasilios A.* Application of Anomaly Detection Algorithms for Detecting SYN Flooding Attacks/ Vasilios A., Fotini Papagalou // IEEE Communication Society Globecom 2004. – 2004. – P. 2050–2054.

10. *Уилер Д.* Статистическое управление процессами: Оптимизация бизнеса с использованием контрольных карт Шухарта./ Д. Уиллер, Д. Чамберс. – М.: Альпина Бизнес Букс, 2009. – 409 с.

11. *Wu X.* A summary of detection of denial-of-QoS attacks on DiffServ networks./X. Wu, V.A. Mahadik, D.S. Reeves// DARPA Information Survivability Conference and Exposition 2003. – April 2003. – P. 277–282.

12. *Hunter J. S.* The Exponentially Weighted Moving Average/ J.S. Hunter// Journal of Quality Technology. –1986. – Vol. 18. – P. 203–209.

13. *Steiner H. S.* Exponentially weighted moving average control charts with time-varying control limits and fast initial response./ Stefan H. Steiner// Journal of Quality Technology. – January 1999. – Vol. 31, № 1. – P. 175–186.

14. EWMA Control Charts // NIST-SEMATECH e-Handbook of Statistical Methods, 2010. URL: http://itl.nist.gov/div898/handbook/pmc/section3/pmc324.htm (дата обращения: 15.02.1012).

15. *Матвеев А.* Байт к байту. Обзор популярных систем учета трафика под Unix/А. Матвеев// Хакер. – Февраль 2009. – С. 30–36.

16. *Tavallaee M. A* Detailed Analysis of the KDD CUP 99 Data Set/ M. Tavallaee, E. Bagheri, W. Lu, Ali A. Ghorbani// Proceedings of the 2009 IEEE Symposium on Computational Intelligence in Security and Defence Applications (CISDA 2009). – 2009. – P. 100–106.

Семенов Николай Александрович – магистрант 1 года обучения кафедры электроники физического факультета ВГУ. E-mail: nick-89@inbox.ru, тел. 89601209311

Телков Александр Юрьевич – доцент кафедры электроники Воронежского государственного университета, кандидат физико-математических наук. E-mail: telkov@dpo-it.ru, тел. 89192438183

Semenov Nikolay Aleksandrovich – First-year student of master's program of chair of Electronics,Voronezh State University. E-mail: nick-89@inbox.ru, tel.89601209311

Telkov Aleksander Yurievich – The Assistant professor of chair of Electronics, Voronezh State University. Candidate of sciences (physics and mathematics). E-mail: telkov@dpo-it.ru, tel. 89192438183