Ранговые распределения для определения пороговых значений сетевых переменных и анализа DDoS атак

А.В. Баскаков[a], Е.С. Сагатов[a], А.М. Сухов[a]

[a] *Самарский национальный исследовательский университет имени академика С.П. Королева, 443086, Московское шоссе, 34, Самара, Россия*

Аннотация

В статье приведен анализ сетевых атак с использованием ранговых распределений. Ранговые распределения для ряда сетевых переменных сравниваются для обычного и аномального сетевых состояний. Подобные исследования были проведены для числа потоков, входящего TCP, UDP трафика, которые генерирует единичный внешний IP адрес. Ранговые распределения, собранные во время обычной эксплуатации сети, позволили определить пороговые значения для сетевых переменных, превышение которых свидетельствует об аномальном поведении сети.

Ключевые слова: пороговые значения сетевых переменных; аномальные сетевые состояния; ранговое распределение для сетевых переменных.

1. Введение

Экспоненциальный рост интернет трафика и числа информационных источников сопровождается быстрым увеличением числа аномальных состояний сети. Аномальные состояния сети объясняются как причинами техногенного характера, так и человеческим фактором. Распознание аномальных состояний, созданных злоумышленниками, достаточно тяжело из-за того, что они имитируют действия обычных пользователей [1]. Поэтому такие аномальные состояния крайне сложно выявить и заблокировать. Задачи обеспечения надёжности и безопасности Интернет сервисов требуют изучения поведения пользователей [2][3] на конкретном ресурсе.

В данной статье пойдёт речь о выявлении аномальных сетевых состояний и методах противодействия DDoS атакам [4][5] (Distributed Denial of Service, распределённая атака типа «отказ в обслуживании») – это такой тип атак, при котором некоторое множество компьютеров в сети Интернет, называемых «зомби», «ботами» или бот сетью (ботнет), по команде злоумышленника начинают отправлять запросы на сервис жертвы. Когда число запросов превышает возможности серверов жертвы, новые запросы от настоящих пользователей перестают обслуживаться и становится недоступным. При этом жертва несёт финансовые убытки.

Исследования, которые описаны в данной статье, используют унифицированный математический подход. Был выделен ряд важнейших сетевых переменных, которые генерирует внешний единичный IP адрес при обращении к заданному серверу или локальной сети. К таким переменным относятся: частота обращения к веб серверу (по заданному порту), число активных потоков, величина входящего TCP, UDP и ICMP трафика и т.д. Построенная инфраструктура позволила измерять величины для вышеперечисленных сетевых переменных.

После нахождения данных величин для анализируемых переменных в произвольный момент времени необходимо построить ранговое распределение. Для этого найденные значения располагаются в порядке убывания. Анализ сетевых состояний будет производиться путем сравнения соответствующих распределений. Особенно наглядно это сравнение, когда распределения для аномального и обычного состояния сети построены на одном графике. Подобный подход позволяет легко определить границу между обычным и аномальным состоянием сети.

Эксперименты по DDoS атаке на сервис можно провести с помощью эмуляции в лабораторных условиях. При этом ценность полученных результатов значительно меньше, чем при DDoS атаке на введённый в эксплуатацию коммерческий сервис, так как эмулятор не может полностью воспроизвести реальную компьютерную сеть. Кроме того, для полноценного понимания принципов и методов DDoS атаки необходим опыт работы с ней. Поэтому авторы анонимно договорились о проведении реальной DDoS атаки на специально подготовленный веб сервис. В процессе атаки был записан сетевой трафик, собрана статистика NetFlow. Изучение ранговых распределений для числа потоков и различных типов входящего трафика, генерируемых единичным внешним IP адресом, позволило определить пороговые значения. Превышение пороговых значений можно классифицировать, как признак атакующего узла, что позволяет сделать выводы об эффективности способов обнаружения и методов противодействия.

2. Обзор предшествующих работ

Идея определения некоторых пороговых значений для выявления аномальных сетевых состояний была высказана достаточно давно, начиная с момента появления DDoS атак [6]. Вопрос состоит только в том, для каких переменных необходимо искать эти пороговые значения. В работе [7] было предложено находить некую статистическую функцию, которая может рассматриваться как энтропия, и по ее поведению делать вывод о начале атаки. Этот подход может быть использован для выявления малозаметных атак с запросами низкой интенсивности [8].

Работа [9] содержит анализ атак на DNS сервера при помощи специально сконструированной величины (detection value) для которой рассчитывается пороговое значение. Эта величина строиться с учетом числа запросов к DNS серверу

и ответов на запросы в течении фиксированного временного интервала. Распознавание атаки с выявлением аномального поведения сети на основе нестандартных отклонений от обычного поведения, то есть когда отклонения значений важнейших сетевых переменных находятся в области превышают три сигмы от нормальных значений описано в статье [10].

Множество работ посвящено обнаружению несанкционированных вторжений при помощи анализа данных протокола NetFlow. Достаточно полный обзор можно найти в статье [11]. Среди методов обнаружения вторжений можно отметить статистический подход [12][13], когда анализ информации о потоках с помощью простейших законов теории вероятности позволяет выявить источники атак. Статистический подход отличается простотой, даёт надежные результаты, но область применения ограничена хорошо изученными типами атак.

Греческие авторы работы [14] предложили использовать для анализа атаки 5 переменных, отношение входящего UDP и ICMP трафика к исходящему, количество потоков, состоящих из одного пакета, количество потоков длительностью меньше 10 миллисекунд, а также количество новых потоков в секунду. Для этих переменных были определены пороговые уровни, при превышении которых можно говорить о начале атаки. Эта работа наиболее близка к предлагаемому нами подходу.

Одна из отличительных особенностей предлагаемого в данной работе подхода это применение рангового анализа для данных NetFlow [15]. Обнаружение атаки базируется на отклонениях от распределения Зипфа [16]. Для этого анализируется информация об активных или завершённых потоках за некоторый промежуток времени. Минимальное время сбора статистики NetFlow может варьироваться от одной до пяти минут. При нормальном функционировании сети ранжированный список количества потоков, генерируемых уникальным IP адресом, представляет типичное распределение Зипфа [17]. Атакующие адреса могут быть идентифицированы по превышению установленного порогового значения для числа активных или завершённых потоков [15]. Во время атак число потоков возрастает многократно.

В настоящее время особое внимание должно быть уделено противостоянию наиболее опасному виду DDoS атак по переполнению внешнего канала, ведущего к отдельному серверу, локальной организации или автономной системе. Следует упомянуть одну из первых аналитических статей [18] по атакам на переполнение внешнего канала, в ней дается обзор источников атаки и стратегий защиты, приведены данные об атакующих мощностях.

3. Ранговые распределения и распознание аномальных сетевых состояний

В 1994 Стив Глассман [19] впервые описал при помощи рангового распределения процесс резервирования интернет трафика. В дальнейшем область применения ранговых распределений для анализа сетевых процессов расширилась, с их помощью были описаны такие интернет процессы, как запросы к поисковым системам, обращения к DNS серверам, популярность документов на сайте и многое другое. В настоящее время доступно несколько хороших обзоров [17][20], посвященных применению ранговых распределений для описания сетевых процессов.

Обычно интернет процессы описываются распределением Зипфа, которое гласит

$$p_i = \frac{p_1}{i^{\alpha}} \tag{1}$$

здесь p_1 - наибольшее значение исследуемой величины, i - порядковый номер в ранжированном списке (списке по убыванию), а α - показатель степени. Следовательно, эти три величины и должны использоваться при анализе атаки.

Для распознания атаки и выявления ее источников сравниваются два ранговых распределения. Одно из этих распределений строиться в текущий момент времени, другое в некоторый предыдущий момент, который рассматривается в качестве нормального состояния сети. Ранее нами было предложено анализировать ранговые распределения для количества потоков, которые генерирует единичный IP адрес [13]. Установлено, что в момент атаки эта величина возрастает не менее, чем на порядок.

То есть для обнаружения момента начала атаки следует использовать величину k

$$k = \frac{p_1}{p_{tr}} \tag{2}$$

Теперь вопрос заключается в том, как определить пороговое значение p_{tr}, которое стоит в знаменателе дроби из уравнения (2).

Для этого следует построить зависимость наибольшего значения исследуемой величины от времени $p_1(t)$. Необходимо собрать и обработать статистику за значительный период времени. Этот период должен составлять не менее недели, чтобы сгладить колебания. О практической реализации этого метода речь пойдет в разделе 6. Здесь же заметим, что на основании этой зависимости можно найти пороговое значение p_{tr}, которое не должно превышаться в процессе нормальной эксплуатации сети $\left(p_1(t) \le p_{tr}\right)$. Именно это значение будет использовано для вычисления коэффициента k из уравнения (2).

Следующий вопрос состоит в определении набора сетевых переменных, для которых необходимо рассчитывать пороговые значения. Выбор сетевых переменных зависит от типа DDoS атаки. Если атака нацелена на нарушение какого-либо интернет сервиса, например, отказ в функционировании веб сервера, то следует анализировать число

запросов к атакуемому ресурсу. Если атака ставит целью переполнение входящих каналов, то требуется собирать данные о всех типах входящего трафика (TCP, UDP, ICMP и др.), а также информацию о числе активных потоков.

Поскольку тип атаки заранее неизвестен, то пороговые значения требуется вычислять для значительного числа переменных. Подобный набор переменных должен включать

- Общее число активных потоков на граничном маршрутизаторе;
- Число активных потоков, которые генерирует единичный внешний IP адрес;
- Входящий трафик, которые генерирует единичный внешний IP адрес, отдельно для каждого типа трафика (TCP, UDP, ICMP и другие);
- Число запросов, которые генерирует единичный внешний IP адрес, отдельно для каждого типа сервиса (HTTP, FTP, mail, proxy, ssh, samba, MySQL, и т.д.).

После того, как пороговые значения p_{tr} для важнейших сетевых переменных будут найдены, необходимо регулярно рассчитывать соответствующие значения коэффициентов, задаваемых уравнением (2). Если значение этого коэффициента значительно превышает единицу, то следует говорить об аномальном состоянии сети.

4. Определение пороговых значений для входящего трафика

В предыдущей работе [13] нами был определен предельный уровень для числа активных потоков, которые генерирует единичный IP адрес. При превышении этого уровня трафик с данного адреса должен быть блокирован на несколько минут.

Проведенная реальная атака показала, что нам необходимо проанализировать входящую скорость основных типов трафика (TCP, UDP, другой трафик), которые генерирует единичный IP адрес. Анализу должны подлежать две величины, первая из них это скорость входящего трафика с каждого из внешних адресов. Вторая зависимость представляет собой недельный график для максимальных скоростей входящего трафика $p_1(t)$.

Для построения искомых графиков были написаны специальные скрипты, которые собирали и анализировали весь входящий трафик, разбивая его по типам трафика. Первая зависимость B_i^{TCP}, представленная на Рис.1, представляет собой ранжированный список IP адресов в логарифмическом масштабе. По оси абсцисс отложен порядковый номер адреса в списке по убыванию, по оси ординат величина входящего трафика.

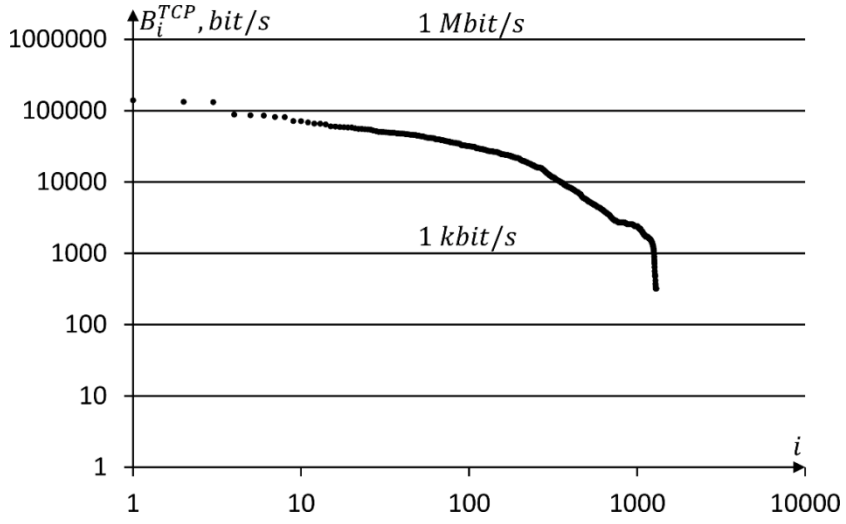

Рис. 1. Ранжированный список для входящего TCP трафика.

Для построения второго графика используется максимальная скорость (количество бит за секунду) с единичного IP адреса B_1^{TCP}, которая достигается на интервале 30 минут. По оси ординат показано время измерения, сбор статистики осуществлялся в течении недели. Зависимость наибольшей скорости с единичного адреса от времени изображена на Рис.2.

Этот график позволяет определить уровень атаки. Он обозначен пунктирной линией. При превышении этого уровня B_{tr}^{TCP}, соответствующий адрес должен быть занесен в подозрительный список и подвергнут процедуре дополнительной проверки.

Наиболее пристальное внимание среди типов трафика следует уделить входящему UDP трафику. Именно он является самым опасным, так как его отправка и приём не требуют подтверждения, и злоумышленник может отправлять его на атакуемый сервер в любых объёмах, тем самым переполняя канал сервера. Однако, при условиях стандартной эксплуатации скорость входящего трафика B_1^{UDP} мала и не менее, чем на порядок уступает уровню TCP трафика.

График, изображенный на Рис. 3, используется для определения уровня вторжения B_{tr}^{UDP}. Следует отметить, что этот уровень во время проведенной нами атаки был превышен не менее, чем на два порядка, или более чем в 100 раз.

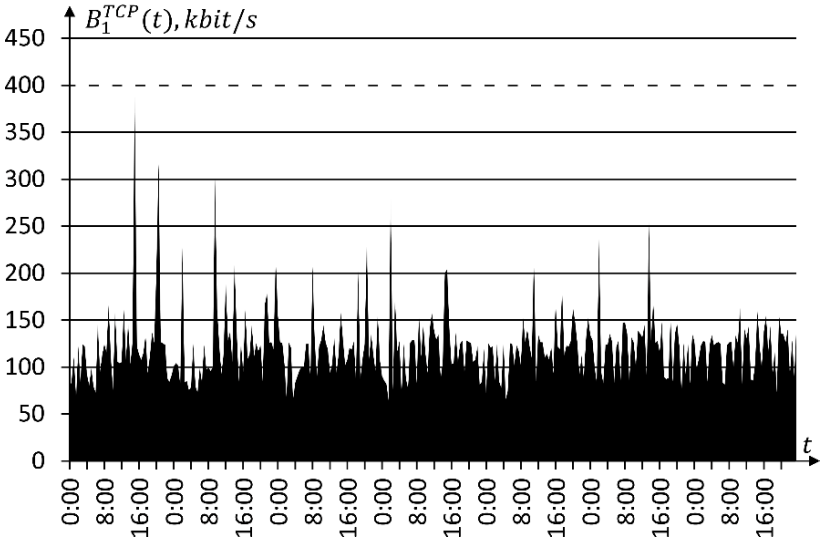

Рис. 2. Зависимость наибольшей входящей скорости TCP потока, генерируемая одиночным адресом.

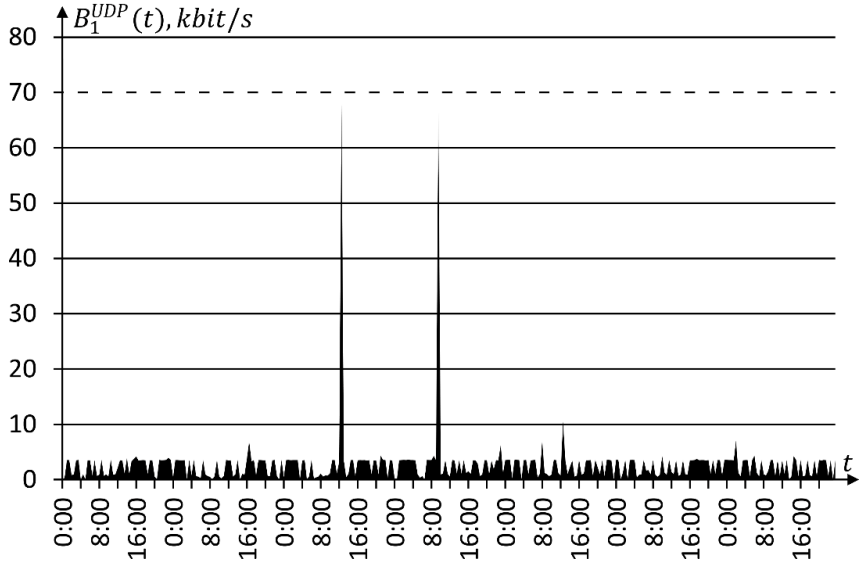

Рис. 3. Зависимость наибольшей входящей скорости UDP потока, генерируемая одиночным адресом.

Так как UDP трафик при DDoS атаках является основной угрозой, то необходимо предусмотреть меры по его быстрой блокировке. Одним из механизмов быстрого оповещения вышестоящих маршрутизаторов может стать использование оповещающих ICMP пакетов. Если UDP порт на компьютере не прослушивается ни одной программой и на него приходит UDP пакет, то обратно отправляется ICMP пакет, говорящий о том, что порт не прослушивается. Для программно-конфигурируемых устройств возможно написание скрипта, который будет блокировать входящий UDP трафик с адреса отправителя на адрес получателя по порту назначения. Возможно также блокирование всего UDP трафика с подозрительного адреса. Главное, чтобы эта блокировка длилась непродолжительное время, 10-15 сек. Существующий функционал iptables легко настроить для отправки таких пакетов.

К сожалению, существующее программное обеспечение маршрутизаторов и коммутаторов не предусматривает возможности анализа ICMP пакетов с последующей блокировкой на время различных типов трафика с подозрительных адресов. Однако применение программно-конфигурируемых сетевых устройств [21] поможет решить эту проблему.

5. Выводы

Для того, чтобы улучшить систему защиты интернет ресурсов было проведено изучение поведения пользователей на примере крупного интернет портала. Поведение пользователя характеризуется рядом сетевых переменных, которые генерирует внешний единичный IP адрес. Для ряда таких переменных как число потоков, входящий TCP, UDP, ICMP трафик, число запросов к различным информационным ресурсам были построены ранговые распределения. На основании зависимости наибольшего значения для этих переменных от времени было найдено пороговое значение, которое характеризует нормальное состояние сети. Превышение этого порогового значения, особенно многократное, является квалификационным признаком аномального сетевого состояния, в том числе и сетевой атаки типа DDoS.

Защита от подобных атак должна осуществляться путем введения ограничений на входящий трафик с атакующих IP адресов, которые определяются по превышению найденного порового значения. Но это ограничение должно быть

установлено не на локальном маршрутизаторе, а у провайдера, как минимум, на маршрутизаторе, что на два уровня выше. Тогда данная защита будет эффективной.

Благодарности

Работа выполнена в рамках государственного задания Министерства образования и науки РФ и при поддержке гранта РФФИ № 16-07-00218а.

Литература

[1] Singh, S. Analysis of Botnet behavior using Queuing theory / Singh S., Gyanchandani M. // International Journal of Computer Science & Communication. – 2010. – ISSN: 0973-7391. – Т. 1. – № 2. – С. 239-241.
[2] Stanton, J. M. Analysis of end user security behaviors / Stanton J. M., Stam K. R., Mastrangelo P., Jolton, J. // Computers & Security. – 2005. – ISSN: 0167-4048. – DOI: 10.1016/j.cose.2004.07.001. – Т. 24. – № 2. – С. 124-133.
[3] Hochheiser, H. Understanding patterns of user visits to web sites: interactive starfield visualizations of WWW log data / Hochheiser, H. Shneiderman, B. // Proceedings of the ASIST Annual Meeting – 1999. – ISSN 0044-7870. – 17с.
[4] Mirkovic, J. A taxonomy of DDoS attack and DDoS defense mechanisms / Mirkovic J., Reiher P. // ACM SIGCOMM Computer Communication Review. – 2004. – ISSN: 0146-4833. – DOI: 10.1145/997150.997156. – Т. 34. – № 2. – С. 39-53.
[5] Douligeris, C. DDoS attacks and defense mechanisms: classification and state-of-the-art / Douligeris C., Mitrokotsa A. // Computer Networks. – 2004. – ISSN: 1389-1286. – DOI: 10.1016/j.comnet.2003.10.003. – Т. 44. – № 5. – С. 643-666.
[6] Jiang, J. Detecting network attacks in the internet via statistical network traffic normality prediction / Jiang J., Papavassiliou S. // Journal of Network and Systems Management. – 2004. – Т. 12. – № 1. – С. 51-72.
[7] Feinstein, L. Statistical approaches to DDoS attack detection and response / Feinstein, L., Schnackenberg, D., Balupari, R., Kindred, D. // DARPA Information Survivability Conference and Exposition, 2003. Proceedings. – IEEE, 2003. – Т. 1. – С. 303-314.
[8] Bhuyan, M.H. Information metrics for low-rate DDoS attack detection: A comparative evaluation / Bhuyan M. H., Bhattacharyya D. K., Kalita J. K. // Contemporary Computing (IC3), 2014 Seventh International Conference on. – IEEE, 2014. – С. 80-84.
[9] Sun, C. Efficient and low-cost hardware defense against DNS amplification attacks / Sun C., Liu B., Shi L. // IEEE GLOBECOM 2008-2008 IEEE Global Telecommunications Conference. – IEEE, 2008. – С. 1-5.
[10] Tan, Z. A system for denial-of-service attack detection based on multivariate correlation analysis / Tan, Z., Jamdagni, A., He, X., Nanda, P., Liu, R. P. // IEEE transactions on parallel and distributed systems. – 2014. – Т. 25. – № 2. – С. 447-456.
[11] Li, B. A survey of network flow applications / Li B., Springer J., Bebis G., Hadi Gunes M. // Journal of Network and Computer Applications. – 2013. – Т. 36. – № 2. – С. 567-581.
[12] François, J. BotTrack: tracking botnets using NetFlow and PageRank / François, J., Wang, S., Engel, T. // NETWORKING 2011. – Springer Berlin Heidelberg, 2011. – С. 1-14.
[13] Sukhov, A.M. Active flows in diagnostic of troubleshooting on backbone links / Sukhov A.M., Sidelnikov D.I., Platonov A.P., Strizhov M.V., Galtsev A.A. // Journal of High Speed Networks. – 2011. – Т. 18. – № 1. – С. 69-81.
[14] Siaterlis, C. Detecting incoming and outgoing DDoS attacks at the edge using a single set of network characteristics / Siaterlis C., Maglaris V. // 10th IEEE Symposium on Computers and Communications (ISCC'05). – IEEE, 2005. – С. 469-475.
[15] Sukhov, A.M. Analysis of Internet service user audiences for network security problems / Sukhov A.M., Sagatov E.S., Baskakov A.V. // Telecommunication Technologies (ISTT), 2014 IEEE 2nd International Symposium on. – IEEE, 2014. – С. 214-219.
[16] Zipf, G.K. Relative frequency as a determinant of phonetic change / Zipf G.K. // Harvard Studies in Classical Philology. – 1929. – С.1-95.
[17] Krashakov, S.A. On the universality of rank distributions of website popularity / Krashakov S.A., Teslyuk A.B., Shchur L.N. // Computer Networks, 2006. – Т. 50. – № 11. – С. 1769-1780.
[18] Geva, M. Bandwidth Distributed Denial of Service: Attacks and Defenses / Geva M., Herzberg A., Gev Y. // IEEE Security & Privacy. – 2014. – Т. 12. – № 1. – С. 54-61.
[19] Glassman, S. A caching relay for the World Wide Web / Glassman S. // Computer Networks and ISDN Systems. – 1994. – Т. 27. – № 2. – С. 165-173.
[20] Dorogovtsev, S.N. Evolution of networks: From biological nets to the Internet and WWW / Dorogovtsev S. N., Mendes J. F. F. // Oxford University Press, 2013. – 263с.
[21] Wickboldt, J.A. Software-defined networking: management requirements and challenges / J.A. Wickboldt, W.P. de Jesus, P.H. Isolani, C.B. Both, J. Rochol, L.Z. Granville //IEEE Communications Magazine. – 2015. – Т. 53. – №. 1. – С. 278-285.