УДК 004.7
MSC 68M25

Автоматическое обнаружение аномалий сетевого трафика при DDoS-атаках

А. В. Орехов[1], *А. А. Орехов*[2]

[1] Санкт-Петербургский государственный университет, Российская Федерация, 199034, Санкт-Петербург, Университетская наб., 7–9
[2] Транстех, Российская Федерация, 196247, Санкт-Петербург, пл. Конституции, 1

Для цитирования: *Орехов А. В., Орехов А. А.* Автоматическое обнаружение аномалий сетевого трафика при DDoS-атаках // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2023. Т. 19. Вып. 2. С. 251–263. https://doi.org/10.21638/11701/spbu10.2023.210

Распределенные атаки типа «отказ в обслуживании» (DDoS-атаки) — это вторжения в вычислительные системы сети Интернет, цель которых — сделать их недоступными для пользователей. DDoS-атаки заключаются в одновременной отправке в сторону определенного ресурса большого количества запросов, в результате чего сервер не выдерживает сетевой нагрузки и доступ к нему становится практически невозможным. В такой ситуации провайдеру необходимо определить момент начала атаки и изменить стратегию управления сетевым трафиком. Обнаружение начала DDoS-атаки возможно методами машинного обучения без учителя, использующими последовательный статистический анализ сетевой активности. Для этого удобно применять математические модели, основанные на дискретных случайных процессах, с монотонно возрастающими траекториями в начале DDoS-атаки. Случайные функции, которые представляют собой соответствие между обобщенным временем и кумулятивным объемом сетевого трафика или между общим количеством входящих пакетов и кумулятивной суммой неотвергнутых пакетов, в начале DDoS-атаки меняют тип своего возрастания с линейного на нелинейный: в первом случае на параболический или экспоненциальный, во втором — на логарифмический или арктангенциальный. Для определения моментов такого изменения в качестве статистических правил можно использовать квадратичные формы аппроксимационно-оценочных критериев.

Ключевые слова: сетевой трафик, DDoS-атака, машинное обучение без учителя, последовательный статистический анализ, марковский момент, метод наименьших квадратов.

1. Введение. Важнейшими задачами любого хостинг-провайдера являются техническое обслуживание серверов, поддержка функциональности соответствующей инфраструктуры, обеспечение бесперебойной работы сайтов и защита данных. В последнее время получили широкое распространение распределенные атаки типа «отказ в обслуживании», или DDoS-атаки, которые применяются для нарушения работы серверной инфраструктуры при помощи отправки огромного числа запросов [1]. Смысл этих сетевых вторжений состоит в том, чтобы подавить работу сервера бóльшим объемом внешнего трафика, с которым тот не может справиться.

Самый распространенный способ организации DDoS-атак — использование для отправки «ложных» запросов так называемых ботнетов, которые состоят из взломанных серверов, компьютеров и других вычислительных устройств, имеющих доступ в Интернет [2, 3], например технологий формирования ботнетов при помощи

взломанных устройств Интернета вещей (IoT) [4]. Жертвой злоумышленников может стать любое устройство IoT, однако чаще всего взлому и захвату подвергаются камеры и роутеры (в силу их распространенности и большого количества), однако в зоне риска находятся медицинское и производственное оборудование, подключенное к IoT, устройства «умного дома» [5, 6]. Кроме этого вида сетевых вторжений различают еще несколько типов DDoS-атак: IP-флуд, SYN-флуд, UDP-флуд, TCP-флуд, Ping of Death, IP-спуфинг, APDoS-атака и т. п. [7, 8]. Для всех перечисленных видов DDoS-атак характерно резкое возрастание объема трафика. Чаще всего нападения происходят через уровни L3, L4 и L7 модели OSI. На L3 и L4 направлены разновидности DDoS-атак — «сетевого» (network layer DDoS) и «транспортного» (transport layer DDoS) уровней. На L7 происходят атаки на прикладном уровне приложений (application layer DDoS) [9]. Если рассматривать уровни L3 и L4, то возрастание сетевой активности при DDoS-атаках проявляется в резком увеличении количества входящих пакетов [10, 11].

Защита от DDoS-атак строится по-разному, в зависимости от многих показателей, например, она зависит от типа хостинга и размещенного на нем ресурса. Аппаратные и программные брандмауэры работают по списку разрешающих и запрещающих правил. Злоумышленники могут нацелиться на открытые порты брандмауэра, которые применяются для доступа легитимных пользователей, поэтому сложные атаки не могут быть обработаны только с помощью стандартного программного брандмауэра. Межсетевые экраны фильтруют трафик на основании четко определяемых списков доступа (правил контроля) и могут регулировать потоки трафика, основываясь на таких критериях как адреса отправителя, используемые сетевые сервисы, порты и протоколы [12]. Из-за относительной легкости исполнения и распределенного характера DDoS-атаки могут быть направлены на сайты любых размеров [13, 14].

Эффективность противодействия DDoS-атакам определяется временем их обнаружения. Ответные действия включают в себя разработку стратегии управления сетевым трафиком, прежде всего изменением маршрутизации, при которой бóльшая часть подозрительных пакетов перенаправляется с сервера, а остальные данные обрабатываются по частям. Остановить DDoS-атаки также позволяют фильтрация входящих данных и разумное отклонение пакетов или соединений, которые могут являться «зловредным» трафиком. Во всех описанных выше случаях провайдер должен применять специализированные аппаратные и программные методы для борьбы с DDoS-атаками [15, 16]. Изменение стратегии управления сетевым трафиком возможно при помощи последовательного статистического анализа [17] и математических моделей, основанных на дискретных случайных процессах.

Случайные функции, которые являются соответствием между обобщенным временем и кумулятивной суммой входящих пакетов или между общим количеством входящих пакетов и кумулятивной суммой неотвергнутых пакетов, по построению монотонно возрастают. При штатных режимах сетевого трафика их рост можно считать «почти линейным». Но в начале DDoS-атаки тип возрастания этих случайных функций изменяется с линейного на нелинейный: в первом случае на параболический или экспоненциальный (рисунок, *а*), во втором — на логарифмический или арктангенциальный (рисунок, *б*). Для определения марковского момента остановки, соответствующего изменению стратегии управления сетевым трафиком в ответ на DDoS-атаку, возможно использование аппроксимационно-оценочных критериев. Основная идея их применения заключается в том, что аппроксимационно-оценочные критерии — это статистические правила в виде квадратичных форм, которые строятся

как разности квадратичной погрешности линейной аппроксимации и квадратичной погрешности нелинейной аппроксимации (в некотором выбранном классе функций) одной и той же числовой последовательности y_t. В момент, когда изменяется характер ее возрастания, квадратичная форма критерия меняет знак, что, в свою очередь, служит достаточным условием для определения соответствующего марковского момента [18]. В качестве случайных величин для выявления аномалий сетевого трафика можно использовать последовательности y_t, которые строятся при помощи счетчиков сетевого экрана, фиксирующих количество входящих или неотвергнутых пакетов на уровнях L3 и L4 модели OSI.

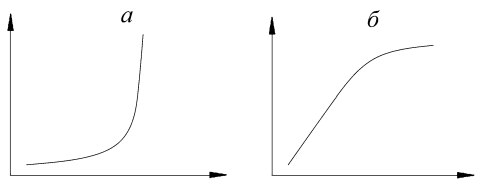

Рисунок. Два эскиза кривой сетевой активности при DDoS-атаках
а — явное задание кривой сетевой активности (на оси абсцисс — обобщенное время, на оси ординат — общее количество входящих пакетов); *б* — параметрическое задание кривой сетевой активности (на оси абсцисс — общее количество входящих пакетов в зависимости от времени, на оси ординат — кумулятивная сумма неотвергнутых пакетов в зависимости от времени).

2. Аппроксимационно-оценочные критерии. Рассмотрим бинарную задачу проверки статистических гипотез H_0 и H_1.

Нулевая гипотеза H_0 — последовательность y_t возрастает линейно, альтернативная гипотеза H_1 — последовательность y_t возрастает нелинейно. Для проверки статистической гипотезы необходимо построить критерий, как строгое математическое правило, позволяющее ее принять или отвергнуть [19]. В общем случае принятие решения в некоторый момент времени может быть основано только на известных значениях дискретного случайного процесса $\xi = \xi(t, \omega)$, где t — дискретное время, ω — случайное событие, принадлежащие некоторому вероятностному пространству $(\Omega, \mathcal{F}, \mathrm{P})$. Если применять формальный подход, то изучаемые события должны быть измеримы в неубывающей последовательности σ-алгебр $\mathfrak{F}_n \in \mathcal{F}$, порожденных процессом $\xi = \xi(t, \omega)$ [20].

Пусть τ — момент наступления некоторого события в случайном процессе $\xi = \xi(t, \omega)$. Если для любого момента времени t_0 можно однозначно сказать, наступило τ или нет при условии, что известны значения процесса $\xi = \xi(t, \omega)$ только в прошлом, то тогда τ — марковский момент относительно неубывающей последовательности σ-алгебр $\mathfrak{F}_n \in \mathcal{F}$, порожденных процессом $\xi = \xi(t, \omega)$ [21, 22]. В рассматриваемом случае марковским моментом остановки случайного процесса $\xi = \xi(t, \omega)$ является минимальное значение τ, при котором отвергается нулевая гипотеза H_0 и принимается альтернативная гипотеза H_1. Для проверки статистических гипотез H_0 и H_1 будем использовать аппроксимационно-оценочные критерии [18].

Квадратичные формы аппроксимационно-оценочных критериев строятся по трем, четырем или пяти узлам аппроксимации. Узлами аппроксимации для числовой последовательности y_t служат упорядоченные пары (i, y_i), где i — натуральный аргумент, y_i — соответствующее значение последовательности y_t. Так как подстрочный

индекс однозначно определяет величину натурального аргумента, для обозначения узла аппроксимации вместо упорядоченной пары (i, y_i) будем использовать элемент последовательности y_i и называть его натуральным узлом аппроксимации.

При построении квадратичных форм аппроксимационно-оценочных критериев применяется прием, который облегчает вычисления. Значения последовательности y_t рассматриваются в узлах $y_0, y_1, \ldots, y_{k-1}$, при этом всегда $y_0 = 0$. Для выполнения такого условия на любом шаге аппроксимации выполняется преобразование:

$$y_0 = y_j - y_j,\ y_1 = y_{j+1} - y_j, \ldots, y_{k-1} = y_{j+k-1} - y_j.$$

Квадратичная погрешность аппроксимации числовой последовательности y_t функцией $f(t)$ в узлах $y_0, y_1, \ldots, y_{k-1}$ равна сумме квадратов разностей y_t и $f(t)$ при соответствующих значениях натурального аргумента:

$$\delta_f^2(k_0) = \sum_{i=0}^{k-1} (f(i) - y_i)^2. \tag{1}$$

Вещественная функция $f(t)$ из некоторого класса Y приближает числовую последовательность y_t по методу наименьших квадратов, если для соответствующей квадратичной формы $\delta_f^2(k_0)$ справедливо выражение

$$\delta_f^2(k_0) = \min_{f \in Y} \sum_{i=0}^{k-1} (f(i) - y_i)^2,$$

такой минимум всегда найдется, так как $\delta_f^2(k_0)$ — положительно определенная квадратичная форма.

Квадратичная погрешность аппроксимации числовой последовательности y_t произвольной нелинейной функцией $f(t)$ по натуральным узлам $y_0, y_1, \ldots, y_{k-1}$ выражается формулой (1), а квадратичная погрешность ее линейной аппроксимации по тем же узлам — формулой

$$\delta_l^2(k_0) = \sum_{i=0}^{k-1} (a \cdot i + b - y_i)^2. \tag{2}$$

В общем случае аппроксимационно-оценочный критерий можно сформулировать следующим образом. Пусть

$$\delta^2(k_0) = \delta_{lf}^2(k_0) = \delta_l^2(k_0) - \delta_f^2(k_0)\,.$$

Будем говорить, что вблизи элемента y_{k-1} тип возрастания y_t изменился с линейного на нелинейный, если для натуральных узлов $y_0, y_1, \ldots, y_{k-1}$ справедливо неравенство $\delta^2(k_0) \leqslant 0$, а для набора узлов $y_1, y_2, \ldots, y_k$, сдвинутых на один шаг дискретности вправо, нелинейная аппроксимация стала точнее линейной, т. е. $\delta^2(k_0) > 0$, иначе, в терминах последовательного статистического анализа. Марковским моментом остановки для случайного процесса $\xi = \xi(t, \omega)$ с монотонно возрастающей траекторией y_t будет

$$\tau = \min\{t \mid\ \delta^2(k_0) > 0\},$$

при котором отвергается гипотеза H_0 и принимается альтернативная гипотеза H_1.

Для обнаружения аномалий сетевого трафика в начале DDoS-атаки можно использовать четыре вида аппроксимационно-оценочных критериев: параболические, экспоненциальные, логарифмические и арктангенциальные [18].

3. Квадратичные формы параболических аппроксимационно-оценочных критериев. Используя метод наименьших квадратов, вычислим коэффициенты a, b линейной функции $f(x) = ax + b$, аппроксимирующей натуральные узлы $y_0, y_1, \ldots, y_{k-1}$. Для этого найдем локальный минимум функции двух переменных:

$$f_l(a,b) = \sum_{i=0}^{k-1} (a \cdot i + b - y_i)^2.$$

Приравняем к нулю ее частные производные по переменным a, b:

$$\frac{\partial f_l}{\partial a} = 2a \sum_{i=0}^{k-1} i^2 + 2b \sum_{i=0}^{k-1} i - 2 \sum_{i=0}^{k-1} i \cdot y_i,$$

$$\frac{\partial f_l}{\partial b} = 2a \sum_{i=0}^{k-1} i + 2b \sum_{i=0}^{k-1} 1 - 2 \sum_{i=0}^{k-1} y_i$$

и решим соответствующую систему линейных уравнений

$$\begin{cases} \dfrac{k(k-1)(2k-1)}{6} \cdot a + \dfrac{k(k-1)}{2} \cdot b = y_1 + 2y_2 + \ldots + (k-1)y_{k-1}, \\ \dfrac{k(k-1)}{2} \cdot a + k \cdot b = y_0 + y_1 + y_2 + \ldots + y_{k-1}. \end{cases}$$

Получим, что

$$a = \frac{6}{k(k^2-1)} \sum_{i=0}^{k-1} (2i + 1 - k) y_i, \quad b = \frac{2}{k(k+1)} \sum_{i=0}^{k-1} (2k - 1 - 3i) y_i. \tag{3}$$

Используя формулы (2) и (3), можно выписать в явном виде линейные аппроксимирующие функции для трех, четырех и пяти натуральных узлов.

Для узлов y_0, y_1, y_2 линейная аппроксимирующая функция имеет вид

$$ax + b = \frac{1}{6}\Big(3y_2 \cdot x + (2y_1 - y_2)\Big).$$

Тогда

$$\delta_l^2(3_0) = \sum_{i=0}^{2} \left(\frac{3y_2 \cdot i + (2y_1 - y_2)}{6} - y_i \right)^2 = \frac{1}{6}\Big(y_2 - 2y_1\Big)^2. \tag{4}$$

Аналогично для узлов y_0, y_1, y_2, y_3 находим, что

$$\delta_l^2(4_0) = \frac{1}{10}\Big(7y_1^2 + 7y_2^2 + 3y_3^2 - 2y_1(2y_2 + y_3) - 8y_2y_3\Big), \tag{5}$$

а для узлов y_0, y_1, y_2, y_3, y_4 —

$$\delta_l^2(5_0) = \frac{1}{10}\Big(7y_1^2 + 8y_2^2 + 7y_3^2 + 4y_4^2 - 2y_1(2y_2 + y_3) - 4y_2(y_3 + y_4) - 8y_3y_4\Big). \tag{6}$$

С помощью метода наименьших квадратов вычислим коэффициенты c, d для неполной параболической функции $cx^2 + d$, аппроксимирующей узлы $y_0, y_1, \ldots, y_{k-1}$. Найдем локальный минимум функции двух переменных:

$$f_q(c, d) = \sum_{i=0}^{k-1}((c \cdot i^2 + d) - y_i)^2.$$

Вычислим частные производные:

$$\frac{\partial f_q}{\partial c} = 2c\sum_{i=0}^{k-1} i^4 + 2d\sum_{i=0}^{k-1} i^2 - 2\sum_{i=0}^{k-1} i^2 \cdot y_i,$$

$$\frac{\partial f_q}{\partial d} = 2c\sum_{i=0}^{k-1} i^2 + 2d\sum_{i=0}^{k-1} 1 - 2\sum_{i=0}^{k-1} y_i,$$

приравняем их к нулю и решим систему из двух линейных уравнений относительно неизвестных c, d:

$$\begin{cases} \dfrac{k(k-1)(2k-1)(3k^2-3k-1)}{30} \cdot c + \dfrac{k(k-1)(2k-1)}{6} \cdot d = \displaystyle\sum_{i=1}^{k-1} i^2 \cdot y_i\,, \\ \dfrac{k(k-1)(2k-1)}{6} \cdot c + k \cdot d = \displaystyle\sum_{i=0}^{k-1} y_i\,, \end{cases}$$

$$c = \frac{30}{k(k-1)(2k-1)(8k^2-3k-11)}\sum_{i=0}^{k-1}(6i^2 - (k-1)(2k-1))y_i, \tag{7}$$

$$d = \frac{6}{k(8k^2-3k-11)}\sum_{i=0}^{k-1}(3k(k-1) - 1 - 5i^2)y_i. \tag{8}$$

Используя формулы (7) и (8), можно выписать в явном виде неполные параболические (без линейного члена) аппроксимирующие функции для натуральных узлов. Вычислим квадратичные погрешности для них, а затем, учитывая соответствующие погрешности линейной аппроксимации (см. формулы (4)–(6)), выпишем в явном виде параболические аппроксимационно-оценочные критерии δ^2_{lq} по трем, четырем и пяти натуральным узлам.

Для узлов y_0, y_1, y_2 получим равенство

$$cx^2 + d = \frac{2}{52}\Big((7y_2 - 2y_1) \cdot x^2 + (12y_1 - 3y_2)\Big).$$

Тогда находим, что

$$\delta^2_q(3_0) = \sum_{i=0}^{2}\left(\frac{2}{52}\Big((7y_2 - 2y_1 \cdot i^2) + (12y_1 - 3y_2)\Big) - y_i\right)^2 = \frac{1}{26}\Big(y_2 - 4y_1\Big)^2.$$

Следовательно,

$$\delta^2_{lq}(3_0) = \delta^2_l(3_0) - \delta^2_q(3_0) = \frac{1}{39}\Big(2y_1^2 - 14y_2y_1 + 5y_2^2\Big).$$

Аналогично для узлов y_0, y_1, y_2, y_3 получим, что

$$\delta_q^2(4_0) = \frac{1}{98}\Big(61y_1^2 + 73y_2^2 + 13y_3^2 - 44y_1y_2 + 6y_1y_3 - 60y_2y_3\Big),$$

$$\delta_{lq}^2(4_0) = \delta_l^2(4_0) - \delta_q^2(4_0) = \frac{1}{245}\Big(19y_1^2 - 11y_2^2 + 41y_3^2 + 12y_1y_2 - 64y_1y_3 - 46y_2y_3\Big),$$

а для узлов y_0, y_1, y_2, y_3, y_4 —

$$\delta_q^2(5_0) = \frac{1}{870}\Big(571y_1^2 + 676y_2^2 + 651y_3^2 + 196y_4^2 - 2y_1(224y_2 + 99y_3 - 76y_4) -$$

$$- 288y_2y_3 - 148y_2y_4 - 648y_3y_4\Big),$$

$$\delta_{lq}^2(5_0) = \delta_l^2(5_0) - \delta_q^2(5_0) = \frac{1}{435}\Big(19y_1^2 + 10y_2^2 - 21y_3^2 + 76y_4^2 +$$

$$+ 2y_1(25y_2 + 6y_3 - 38y_4) - 10y_2(3y_3 + 10y_4) - 24y_3y_4\Big).$$

4. Аппроксимационно-оценочные критерии с иррациональными коэффициентами. Построим аппроксимационно-оценочные критерии для трех классов нелинейных функций: экспоненциальных — $pe^x + q$, логарифмических — $g\ln(x + 1) + h$ и арктангенсов — $w\arctan x + v$. В общем случае для этих функций коэффициенты соответствующих квадратичных форм являются иррациональными числами. Поэтому в отличие от параболических аппроксимационно-оценочных критериев коэффициенты экспоненциальных, логарифмических и арктангенциальных аппроксимационно-оценочных критериев можно вычислить только приближенно.

Легко заметить, что все три аппроксимирующие функции имеют одинаковую структуру относительно неизвестных коэффициентов: $\alpha\varphi(x) + \beta$. Используя метод наименьших квадратов, вычислим коэффициенты α и β.

Найдем локальный минимум функции двух переменных:

$$f(\alpha, \beta) = \sum_{i=0}^{k-1}(\alpha\varphi(i) + \beta - y_i)^2.$$

Сначала вычислим частные производные функции $f(\alpha, \beta)$ и приравняем их к нулю:

$$\frac{\partial f}{\partial \alpha} = 2\sum_{i=0}^{k-1}\varphi(i)(\alpha\varphi(i) + \beta - y_i),$$

$$\frac{\partial f_e}{\partial \beta} = 2\sum_{i=0}^{k-1}(\alpha\varphi(i) + \beta - y_i),$$

$$\begin{cases} \alpha \cdot \sum\limits_{i=0}^{k-1}\varphi(i)^2 + \beta \cdot \sum\limits_{i=0}^{k-1}\varphi(i) = \sum\limits_{i=1}^{k-1}\varphi(i)y_i, \\ \alpha \cdot \sum\limits_{i=0}^{k-1}\varphi(i) + k\beta = \sum\limits_{i=1}^{k-1} y_i. \end{cases}$$

Затем решим систему линейных уравнений относительно неизвестных α и β:

$$\alpha = \frac{k \cdot \sum_{i=1}^{k-1} \varphi(i) y_i - \sum_{i=0}^{k-1} \varphi(i) \cdot \sum_{i=1}^{k-1} y_i}{k \cdot \sum_{i=0}^{k-1} \varphi(i)^2 - \left(\sum_{i=0}^{k-1} \varphi(i)\right)^2}, \tag{9}$$

$$\beta = \frac{\sum_{i=1}^{k-1} y_i \cdot \sum_{i=0}^{k-1} \varphi(i)^2 - \sum_{i=0}^{k-1} \varphi(i) \cdot \sum_{i=1}^{k-1} \varphi(i) y_i}{k \cdot \sum_{i=0}^{k-1} \varphi(i)^2 - \left(\sum_{i=0}^{k-1} \varphi(i)\right)^2}\ . \tag{10}$$

Квадратичная погрешность аппроксимации натуральных узлов в классе экспоненциальных функций вида $pe^x + q$ равна $\quad \delta_e^2(k_0) = \sum\limits_{i=0}^{k-1} (pe^i + q - y_i)^2$.

Используя формулы (9) и (10), получим, что

$$p = \frac{k \cdot \sum_{i=1}^{k-1} e^i y_i - \sum_{i=0}^{k-1} e^i \cdot \sum_{i=1}^{k-1} y_i}{k \cdot \sum_{i=0}^{k-1} e^{2i} - \left(\sum_{i=0}^{k-1} e^i\right)^2},$$

$$q = \frac{\sum_{i=1}^{k-1} y_i \cdot \sum_{i=0}^{k-1} e^{2i} - \sum_{i=0}^{k-1} e^i \cdot \sum_{i=1}^{k-1} e^i y_i}{k \cdot \sum_{i=0}^{k-1} e^{2i} - \left(\sum_{i=0}^{k-1} e^i\right)^2}\ .$$

Аналогично построению параболических аппроксимационно-оценочных критериев вычислим коэффициенты квадратичных форм для экспоненциальных, логарифмических и арктангенциальных критериев.

Для экспоненциального критерия по трем узлам y_0, y_1, y_2 находим, что

$$\delta_e^2(3_0) \simeq 0.6224 y_1^2 - 0.33476 y_1 y_2 + 0.045015 y_2^2,$$

$$\delta_{le}^2(3_0) = \delta_l^2(3_0) - \delta_e^2(3_0) \simeq 0.044302 y_1^2 - 0.33191 y_1 y_2 + 0.12165 y_2^2,$$

по натуральным узлам y_0, y_1, y_2, y_3 —

$$\delta_e^2(4_0) \simeq 0.6344 y_1^2 + 0.749 y_2^2 + y_1 \left(-0.5186 y_2 + 0.05939 y_3\right) - 0.4549 y_2 y_3 + 0.0735 y_3^2,$$

$$\delta_{le}^2(4_0) = \delta_l^2(4_0) - \delta_e^2(4_0) \simeq 0.06563 y_1^2 - 0.04925 y_2^2 + \\ + y_1 \left(0.1186 y_2 - 0.2594 y_3\right) - 0.3451 y_2 y_3 + 0.2265 y_3^2$$

и для узлов y_0, y_1, y_2, y_3, y_4 —

$$\delta_e^2(5_0) \simeq 0.694 y_1^2 + 0.752 y_2^2 + 0.796 y_3^2 + y_2 \left(-0.371 y_3 - 0.02968 y_4\right) + \\ + y_1 \left(-0.543 y_2 - 0.357 y_3 + 0.1474 y_4\right) - 0.511 y_3 y_4 + 0.0904 y_4^2,$$

$$\delta_{le}^2(5_0) = \delta_l^2(5_0) - \delta_e^2(5_0) \simeq 0.00556 y_1^2 + 0.0483 y_2^2 - 0.0957 y_3^2 + \\ + y_2 \left(-0.02895 y_3 - 0.370 y_4\right) + y_1 \left(0.1428 y_2 + 0.1572 y_3 - 0.1474 y_4\right) - 0.2890 y_3 y_4 + 0.3096 y_4^2.$$

Квадратичная погрешность аппроксимации натуральных узлов в классе логарифмических функций вида $g \ln(x+1) + h$ равна $\quad \delta_n^2(k_0) = \sum\limits_{i=0}^{k-1} (g \ln(i+1) + h - y_i)^2$.

Используя формулы (9) и (10), получим, что

$$g = \frac{k \cdot \sum_{i=1}^{k-1} \ln(i+1) y_i - \sum_{i=0}^{k-1} \ln(i+1) \cdot \sum_{i=1}^{k-1} y_i}{k \cdot \sum_{i=0}^{k-1} \ln^2(i+1) - \left(\sum_{i=0}^{k-1} \ln(i+1)\right)^2},$$

$$h = \frac{\sum_{i=1}^{k-1} y_i \cdot \sum_{i=0}^{k-1} \ln^2(i+1) - \sum_{i=0}^{k-1} \ln(i+1) \cdot \sum_{i=1}^{k-1} \ln(i+1) y_i}{k \cdot \sum_{i=0}^{k-1} \ln^2(i+1) - \left(\sum_{i=0}^{k-1} \ln(i+1)\right)^2}.$$

Для узлов y_0, y_1, y_2 справедливы равенства

$$\delta_n^2(3_0) \simeq 0.65177 y_1^2 - 0.82244 y_1 y_2 + 0.25945 y_2^2,$$

$$\delta_{ln}^2(3_0) = \delta_l^2(3_0) - \delta_n^2(3_0) \simeq 0.0148974 y_1^2 + 0.155775 y_1 y_2 - 0.092785 y_2^2,$$

для узлов y_0, y_1, y_2, y_3 —

$$\delta_n^2(4_0) \simeq 0.74052 y_1^2 + 0.66471 y_2^2 + y_1 (-0.44314 y_2 - 0.38934 y_3) - 0.83197 y_2 y_3 + 0.42699 y_3^2,$$

$$\delta_{ln}^2(4_0) = \delta_l^2(4_0) - \delta_n^2(4_0) \simeq -0.040523 y_1^2 + 0.035294 y_2^2 + \\ + y_1 (0.043138 y_2 + 0.18934 y_3) + 0.031966 y_2 y_3 - 0.12699 y_3^2$$

и для узлов y_0, y_1, y_2, y_3, y_4 —

$$\delta_n^2(5_0) \simeq 0.75674 y_1^2 + 0.78767 y_2^2 + 0.68619 y_3^2 + 0.53691 y_4^2 + \\ + y_2 (-0.47491 y_3 - 0.51389 y_4) + y_1 (-0.35382 y_2 - 0.25967 y_3 - 0.18664 y_4) - 0.74609 y_3 y_4,$$

$$\delta_{ln}^2(5_0) = \delta_l^2(5_0) - \delta_n^2(5_0) \simeq -0.056743 y_1^2 + 0.0123264 y_2^2 + 0.0138145 y_3^2 - 0.136906 y_4^2 + \\ + y_2 (0.074911 y_3 + 0.113895 y_4) + y_1 (-0.046182 y_2 + 0.059668 y_3 + 0.18664 y_4) - 0.053914 y_3 y_4.$$

Квадратичная погрешность аппроксимации натуральных узлов в классе функций вида $f(x) = w \arctan x + v$ равна $\delta_a^2(k_0) = \sum_{i=0}^{k-1} (w \arctan i + v - y_i)^2$.

Используя формулы (9) и (10), получим, что

$$w = \frac{k \cdot \sum_{i=1}^{k-1} \arctan i \cdot y_i - \sum_{i=0}^{k-1} \arctan i \cdot \sum_{i=1}^{k-1} y_i}{k \cdot \sum_{i=0}^{k-1} \arctan^2 i - \left(\sum_{i=0}^{k-1} \arctan i\right)^2},$$

$$v = \frac{\sum_{i=1}^{k-1} y_i \cdot \sum_{i=0}^{k-1} \arctan^2 i - \sum_{i=0}^{k-1} \arctan i \cdot \sum_{i=1}^{k-1} \arctan i \cdot y_i}{k \cdot \sum_{i=0}^{k-1} \arctan^2 i - \left(\sum_{i=0}^{k-1} \arctan i\right)^2}.$$

Для узлов y_0, y_1, y_2 находим, что

$$\delta_a^2(3_0) \simeq 0.62985 y_1^2 - 0.89361 y_1 y_2 + 0.31696 y_2^2,$$

$$\delta_{la}^2(3_0) = \delta_l^2(3_0) - \delta_a^2(3_0) \simeq 0.036820 y_1^2 + 0.226946 y_1 y_2 - 0.150292 y_2^2,$$

для натуральных узлов y_0, y_1, y_2, y_3 —

$$\delta_a^2(4_0) \simeq 0.75 y_1^2 + 0.63932 y_2^2 + y_1 (-0.5 y_2 - 0.5 y_3) - 0.81898 y_2 y_3 + 0.52017 y_3^2,$$

$$\delta_{la}^2(4_0) = \delta_l^2(4_0) - \delta_a^2(4_0) \simeq -0.05 y_1^2 + 0.06068 y_2^2 + \\ + y_1 (0.1 y_2 + 0.3 y_3) + 0.01898 y_2 y_3 - 0.22017 y_3^2$$

и для узлов y_0, y_1, y_2, y_3, y_4 —

$$\delta_a^2(5_0) \simeq 0.79001y_1^2 + 0.76095y_2^2 + 0.69185y_3^2 + y_2\,(-0.52998y_3 - 0.55804y_4) +$$
$$+ y_1\,(-0.36049y_2 - 0.33425y_3 - 0.32005y_4) - 0.66300y_3y_4 + 0.64011y_4^2,$$
$$\delta_{la}^2(5_0) = \delta_l^2(5_0) - \delta_a^2(5_0) \simeq -0.090007y_1^2 + 0.039054y_2^2 + 0.0081498y_3^2 - 0.24011y_4^2 +$$
$$+ y_2\,(0.129980y_3 + 0.15804y_4) +$$
$$+ y_1\,(-0.039511y_2 + 0.134249y_3 + 0.32005y_4) - 0.136998y_3y_4.$$

5. Заключение. Гибкость набора аппроксимационно-оценочных критериев позволяет применять их для принятия решений по смене стратегии управления сетевым трафиком на основе любой доступной метрики. Контроль за загруженностью сети можно рассматривать как систему с переключениями [23, 24].

Формирование нового набора натуральных узлов $y_{t_0-k}, \ldots, y_{t_0-2}, y_{t_0-1}$ из левой полуокрестности точки y_{t_0} — случайное событие $\mathbf{\Omega_{t_0}}$; ему будет соответствовать определенное значение квадратичной формы некоторого аппроксимационно-оценочного критерия, которое обозначим как $\delta_{t_0}^2$.

Рассмотрим последовательность случайных событий:

$$\mathbf{\Omega_k}, \ldots, \mathbf{\Omega_t}, \ldots, \tag{11}$$

и поставим им в соответствие двухэлементное множество исходов $\{C, B\}$, где исход C — событие $\delta_{t_0}^2 \leqslant 0$ и B — событие $\delta_{t_0}^2 > 0$. Так как вероятность наступления либо C, либо B зависит только от набора $y_{t_0-k}, \ldots, y_{t_0-2}, y_{t_0-1}$, то последовательность случайных событий (11) является цепью Маркова с памятью порядка k [25].

В этом случае принятие решения об изменении стратегии управления сетевым трафиком производится за счет марковского момента остановки, который можно определить при помощи аппроксимационно-оценочного критерия. Очевидно, что в такой ситуации переключения могут осуществляться без участия человека, а управляющий субъект имеет аппаратную реализацию.

Предложенный математический аппарат может стать основой для разработки доступных и эффективных аппаратно-программных решений по обеспечению защиты инфраструктурных элементов сети Интернет от DDoS-атак. На основе изложенного материала могут быть реализованы системы, в том числе и с открытым исходным кодом, применимые на практике и обладающие свойством прозрачности и доказательности в противовес развивающимся проприетарным решениям, алгоритмы работы которых являются коммерческой тайной. К важным аспектам рассмотренных инструментов относится то, что решение о смене стратегии принимается на основе данных об изменении характера сетевой активности, а не на основании преодоления неких фиксированных значений нагрузки. Это позволяет применять предложенные методы без дополнительной настройки на системах любого масштаба.

Литература

1. *Gu Q., Liu P.* Denial of service attacks // Handbook of Computer Networks. Hoboken, New Jersey: John Wiley and Sons, 2012. Vol. 3. P. 454–468. https://doi.org/10.1002/9781118256107.ch29

2. *Burghouwt P., Spruit M., Sips H.* Towards detection of botnet communication through social media by monitoring user activity // Information systems security / eds by S. Jajodia, C. Mazumdar. ICISS 2011. Lecture Notes in Computer Science. Vol. 7093. Berlin; Heidelberg: Springer, 2011. P. 131–143. https://doi.org/10.1007/978-3-642-25560-1_9

3. *Schiller C. A., Binkley J., Harley D., Evron G., Bradley T., Willems C., Cross M.* Botnets: The Killer Web Applications. 1st ed. Burlington, Virginia: Syngress, February 15, 2007. 480 p.

4. *Dzaferovic E., Sokol A., Almisreb A. A., Norzeli A. S. M.* DoS and DDoS vulnerability of IoT: A review // Sustainable Engineering and Innovation. 2019. Vol. 1(1). P. 43–48. https://doi.org/10.37868/sei.v1i1.36

5. *Alieyan K., Almomani A., Abdullah R., Almutairi B., Alauthman M.* Botnet and Internet of Things (IoTs): A definition, taxonomy, challenges, and future directions // Security, privacy, and forensics issues in big data / eds by R. Joshi, B. Gupta. Hershney, PA: IGI Global, 2020. P. 304–316. https://doi.org/10.4018/978-1-5225-9742-1.ch013

6. *Dange S., Chatterjee M.* IoT Botnet: The largest threat to the IoT network // Data Communication and Networks. Advances in Intelligent Systems and Computing / eds by L. Jain, G. Tsihrintzis, V. Balas, D. Sharma. Singapore: Springer, 2020. Vol. 1049. P. 137–157. https://doi.org/10.1007/978-981-15-0132-6_10

7. *Alhammadi N. A. M., Zaboon K. H., Abdullah A. A.* A review of the common DDoS attack: types and protection approaches based on artificial intelligence // Fusion: Practice and Applications, 2022. Vol. 7. N 1. P. 8–14. https://doi.org/10.54216/FPA.070101

8. *Бекенева Я. А.* Анализ актуальных типов DDoS-атак и методов защиты от них // Известия СПбГЭТУ «ЛЭТИ». 2016. № 1. С. 7–14.

9. *Obaid H. S., Abeed E. H.* DoS and DDoS attacks at OSI layers // International Journal of Multidisciplinary Research and Publications (IJMRAP). 2020. Vol. 2. Iss. 8. P. 1–9.

10. *Alashhab Z. R., Anbar M., Singh M. M., Hasbullah I. H., Jain P., Al-Amiedy T. A.* Distributed denial of service attacks against cloud computing environment: survey, issues, challenges and coherent taxonomy // Appl. Sci. 2022. Vol. 12. N 12441. https://doi.org/10.3390/app122312441

11. *Kleyman B.* Why DDoS is more dangerous for cloud and data center providers. February 9, 2023. URL: https://www.datacenterfrontier.com/sponsored/article/21545878/a10-why-ddos-is-more-dangerous-for-cloud-and-data-center-providers (дата обращения: 20.02.2023)

12. *Евглевская Н. В., Зуев А. Ю., Карасенко А. О., Лаута О. С.* Сравнительный анализ эффективности существующих методов защиты сетей связи от DDoS-атак // Радиопромышленность. 2020. Т. 30. № 3. С. 67–74. https://doi.org/10.21778/2413-9599-2020-30-3-67-74

13. *Aamir M., Zaidi M. A.* A survey on DDoS attack and defense strategies: from traditional schemes to current techniques // Interdisciplinary Information Sciences. 2013. Vol. 19(2). P. 173–200. https://doi.org/10.4036/iis.2013.173

14. *Mahajan D., Sachdeva M.* DDoS attack prevention and mitigation techniques — a review // International Journal of Computer Applications. April 2013. Vol. 67(19). P. 21–24. https://doi.org/10.5120/11504-7221

15. *Rustam F., Mushtaq M. F., Hamza A., Farooq M. S., Jurcut A. D., Ashraf I.* Denial of service attack classification using machine learning with multi-features // Electronics. 2022. Vol. 11. P. 3817. https://doi.org/10.3390/electronics11223817

16. *Ahmed S., Khan Z. A., Mohsin S. M., Latif S., Aslam S., Mujlid H., Adil M., Najam Z.* Effective and efficient DDoS attack detection using Deep Learning algorithm, multi-layer perceptron // Future Internet. 2023. Vol. 15. N 76. https://doi.org/10.3390/fi15020076

17. *Wald A.* Sequential analysis. New York, USA: John Wiley & Sons, 1947. 212 p.

18. *Orekhov A. V.* Quasi-deterministic processes with monotonic trajectories and unsupervised machine learning // Mathematics. 2021. Vol. 9. N 2301. https://doi.org/10.3390/ math9182301

19. *Lehmann E. L., Romano J. P.* Testing statistical hypotheses. New York: Springer-Verlag, 2005. N XIV. 786 p.

20. *Мазалов В. В.* Математическая теория игр и приложения. СПб.: Лань, 2017. 448 с.

21. *Булинский А. В., Ширяев А. Н.* Теория случайных процессов. М.: Физматлит, Лаборатория базовых знаний, 2003. 400 с.

22. *Shiryaev A. N.* Optimal stopping rules. Berlin; Heidelberg: Springer-Verlag, 2008. N XII. 220 p. https://doi.org/10.1007/978-3-540-74011-7

23. *Shorten R., Wirth F., Mason O., Wulff K., King C.* Stability criteria for switched and hybrid systems // SIAM Review. 2007. Vol. 49. N 4. P. 545–592. https://doi.org/10.1137/05063516X

24. *Hespanha J. P.* Stochastic hybrid systems: application to communication networks. Hybrid systems: Computation and Control. HSCC 2004. Lecture Notes in Computer Science / eds by R. Alur, G. J. Pappas. Berlin; Heidelberg: Springer, 2004. Vol. 2993. P. 387–401. https://doi.org/10.1007/978-3-540-24743-2_26

25. *Wu Sh.-J., Chu M. T.* Markov chains with memory, tensor formulation, and the dynamics of power iteration // Applied Mathematics and Computation. 2017. Vol. 303. P. 226–239. https://doi.org/10.1016/j.amc.2017.01.030

Статья поступила в редакцию 25 февраля 2023 г.
Статья принята к печати 25 апреля 2023 г.

Контактная информация:

Орехов Андрей Владимирович — ст. преп.; a_v_orehov@mail.ru

Орехов Алексей Андреевич — opexob@yandex.ru

Network traffic anomalies automatic detection in DDoS attacks

A. V. Orekhov[1], *A. A. Orekhov*[2]

[1] St. Petersburg State University, 7–9, Universitetskaya nab., St. Petersburg, 199034, Russian Federation
[2] Transtech, 1, pl. Konstitutsii, St. Petersburg, 196247, Russian Federation

For citation: Orekhov A. V., Orekhov A. A. Network traffic anomalies automatic detection in DDoS attacks. *Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes*, 2023, vol. 19, iss. 2, pp. 251–263.
https://doi.org/10.21638/11701/spbu10.2023.210 (In Russian)

Distributed denial-of-service attacks (DDoS attacks) are intrusions into computing systems of the Internet. Their purpose is to make systems of the Internet inaccessible for users. DDoS attack consist of sending many requests to a certain resource at the same time. As a result, the server cannot withstand the network load. In such situation, a provider must determine the moment when attack begins and change the traffic management strategy. Detection of the beginning of a DDoS attack is possible by using unsupervised machine learning methods and sequential statistical analysis of network activity. To activate that, convenient to use mathematical models based on discrete random processes with monotonically increasing trajectories. Random functions, which are represented in the correspondence between generalized time and the cumulative sum of network traffic or the correspondence between the total number of incoming packets and the cumulative sum of packets processed, change their type of increasing from linear to non-linear. In the first case, to parabolic or exponential, in the second case to logarithmic or arctangent. To determine the moment when the type of increasing is going to change, one can use quadratic forms of approximation-estimation tests as statistical rules.

Keywords: traffic strategy, DDoS attack, unsupervised machine learning, sequential statistical analysis, Markov moment, least squares method.

References

1. Gu Q., Liu P. Denial of service attacks. *Handbook of Computer Networks.* Hoboken, New Jersey, John Wiley and Sons Publ., 2012, vol. 3, pp. 454–468. https://doi.org/10.1002/9781118256107.ch29

2. Burghouwt P., Spruit M., Sips H. Towards detection of botnet communication through social media by monitoring user activity. *Information systems security. Eds by S. Jajodia, C. Mazumdar. ICISS 2011. Lecture Notes in Computer Science. Vol. 7093.* Berlin, Heidelberg, Springer Publ., 2011, pp. 131–143. https://doi.org/10.1007/978-3-642-25560-1_9

3. Schiller C. A., Binkley J., Harley D., Evron G., Bradley T., Willems C., Cross M. *Botnets: The Killer Web Applications.* 1st ed. Burlington, Virginia, Syngress Publ., February 15, 2007, 480 p.

4. Dzaferovic E., Sokol A., Almisreb A. A., Norzeli A. S. M. DoS and DDoS vulnerability of IoT: A review. *Sustainable Engineering and Innovation*, 2019, vol. 1(1), pp. 43–48.
https://doi.org/10.37868/sei.v1i1.36

5. Alieyan K., Almomani A., Abdullah R., Almutairi B., Alauthman M. Botnet and Internet of Things (IoTs): A definition, taxonomy, challenges, and future directions. *Security, privacy, and forensics issues in big data.* Eds by R. Joshi, B. Gupta. Hershney, PA, IGI Global Publ., 2020, pp. 304–316.
https://doi.org/10.4018/978-1-5225-9742-1.ch013

6. Dange S., Chatterjee M. IoT Botnet: The largest threat to the IoT network. *Data Communication and Networks. Advances in Intelligent Systems and Computing.* Eds by L. Jain, G. Tsihrintzis, V. Balas, D. Sharma. Singapore, Springer Publ., 2020, vol. 1049, pp. 137–157.
https://doi.org/10.1007/978-981-15-0132-6_10

7. Alhammadi N. A. M., Zaboon K. H., Abdullah A. A. A review of the common DDoS attack: types and protection approaches based on artificial intelligence. *Fusion: Practice and Applications*, 2022, vol. 7, no. 1, pp. 8–14. https://doi.org/10.54216/FPA.070101

8. Bekeneva Ya. A. Analiz aktual'nykh tipov DDoS-atak i metodov zashchity ot nikh [Analysis of actual types of DDoS attacks and methods of protection against them]. *Proceedings of St. Petersburg Electrotechnical University "LETI"*, 2016, no. 1, pp. 7–14. (In Russian)

9. Obaid H. S., Abeed E. H. DoS and DDoS attacks at OSI layers. *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, 2020, vol. 2, iss. 8, pp. 1–9.

10. Alashhab Z. R., Anbar M., Singh M. M., Hasbullah I. H., Jain P., Al-Amiedy T. A. Distributed denial of service attacks against cloud computing environment: survey, issues, challenges and coherent taxonomy. *Appl. Sci.*, 2022, vol. 12, no. 12441. https://doi.org/10.3390/app122312441

11. Kleyman B. *Why DDoS is more dangerous for cloud and data center providers.* February 9, 2023. Available at: https://www.datacenterfrontier.com/sponsored/article/21545878/a10-why-ddos-is-more-dangerous-for-cloud-and-data-center-providers (accessed: February 20, 2023).

12. Evglevskaya N. V., Zuev A. Yu., Karasenko A. O., Lauta O. S. Sravnitel'nyi analiz effektivnosti sushchestvuiushchikh metodov zashchity setei sviazi ot DDoS atak [Comparative analysis of the effectiveness of existing methods of networks security from DDoS attacks]. *Radio industry*, 2020, vol. 30, no. 3, pp. 67–74. https://doi.org/10.21778/2413-9599-2020-30-3-67-74 (In Russian)

13. Aamir M., Zaidi M. A. A survey on DDoS attack and defense strategies: from traditional schemes to current techniques. *Interdisciplinary Information Sciences*, 2013, vol. 19(2), pp. 173–200. https://doi.org/10.4036/iis.2013.173

14. Mahajan D., Sachdeva M. DDoS attack prevention and mitigation techniques — a review. *International Journal of Computer Applications*, April 2013, vol. 67(19), pp. 21–24. https://doi.org/10.5120/11504-7221

15. Rustam F., Mushtaq M. F., Hamza A., Farooq M. S., Jurcut A. D., Ashraf I. Denial of service attack classification using machine learning with multi-features. *Electronics*, 2022, vol. 11, no. 3817. https://doi.org/10.3390/electronics11223817

16. Ahmed S., Khan Z. A., Mohsin S. M., Latif S., Aslam S., Mujlid H., Adil M., Najam Z. Effective and efficient DDoS attack detection using Deep Learning algorithm, multi-layer perceptron. *Future Internet*, 2023, vol. 15, no. 76. https://doi.org/10.3390/fi15020076

17. Wald A. *Sequential Analysis.* New York, USA, John Wiley & Sons Publ., 1947, 212 p.

18. Orekhov A. V. Quasi-deterministic processes with monotonic trajectories and unsupervised machine learning. *Mathematics*, 2021, vol. 9, no. 2301. https://doi.org/10.3390/ math9182301

19. Lehmann E. L., Romano J. P. *Testing statistical hypotheses.* New York, Springer-Verlag Publ., 2005, no. XIV, 786 p.

20. Mazalov V. V. *Matematicheskaia teoriia igr i prilozheniia* [*Mathematical game theory and applications*]. St. Petersburg, Lan' Publ., 2017, 448 p. (In Russian)

21. Bulinsky A. V., Shiryaev A. N. *Teoriia sluchaynykh protsessov* [*Theory of random processes*]. Moscow, Fizmatlit Laboratory of basic knowledge Publ., 2003, 400 p. (In Russian)

22. Shiryaev A. N. *Optimal stopping rules.* Berlin, Heidelberg, Springer-Verlag Publ., 2008, no. XII, 220 p. https://doi.org/10.1007/978-3-540-74011-7

23. Shorten R., Wirth F., Mason O., Wulff K., King C. Stability criteria for switched and hybrid systems. *SIAM Review*, 2007, vol. 49, no. 4, pp. 545–592. https://doi.org/10.1137/05063516X

24. Hespanha J. P. Stochastic hybrid systems: application to communication networks. *Hybrid Systems: Computation and Control. HSCC 2004. Lecture Notes in Computer Science.* Eds by R. Alur, G. J. Pappas. Berlin, Heidelberg, Springer Publ., 2004, vol. 2993, pp. 387–401. https://doi.org/10.1007/978-3-540-24743-2_26

25. Wu Sh.-J., Chu M. T. Markov chains with memory, tensor formulation, and the dynamics of power iteration. *Applied Mathematics and Computation*, 2017, vol. 303, pp. 226–239. https://doi.org/10.1016/j.amc.2017.01.030

Received: February 25, 2023.
Accepted: April 25, 2023.

Authors' information:

Andrey V. Orekhov — Senior Lecturer; a_v_orehov@mail.ru

Aleksey A. Orekhov — opexob@yandex.ru