DOI: 10.55648/1998-6920-2023-17-2-44-50 УДК 004.056

Обзор методов прогнозирования сетевых аномалий

Д. С. Лизнев

Сибирский гос. унив. телекоммуникаций и информатики (СибГУТИ)

Аннотация: В работе проведен анализ методов прогнозирования сетевых аномалий. На примере реальных статистических данных показаны этапы настройки моделей прогнозирования. Показано влияние DDoS-атак на энтропию IP-адресов назначения.

Ключевые слова: модель экспоненциального сглаживания, авторегрессионная модель, энтропия, сетевые атаки.

Для цитирования: Лизнев Д. С. Обзор методов прогнозирования сетевых аномалий // Вестник СибГУТИ. 2023. Т. 17, № 2. С. 44–50. https://doi.org/10.55648/1998-6920-2023-17-2-44-50.

Статья поступила в редакцию 26.12.2022;
принята к публикации 10.01.2023.

1. Введение

Сетевая аномалия – действия с применением программных и (или) технических средств и с использованием сетевого протокола, направленные на реализацию угроз несанкционированного доступа к информации, воздействия на нее или на ресурсы автоматизированной системы [1].

Атаки можно классифицировать как Denial of Service/Distributed Denial of Service (DoS/DDoS), User to Root (U2R), Remote to Local (R2L) и Probe. В данной статье рассматривается возможность применения статистических методов для прогнозирования количества DDoS-атак, а также влияние указанного класса атаки на энтропию IP-адресов назначения.

DDoS-атаки приводят к тому, что целевая система становится полностью недоступной или нестабильной. Данный класс атаки делят на несколько типов, например, такие как Smurf (рассылка поддельных ICMP-запросов), Land (рассылка некорректных TCP-запросов), Neptune (одновременная множественная рассылка SYN-сегментов TCP) и т.д.

Согласно отчету [2] количество проводимых DDoS-атак неуклонно возрастает (рис. 1).

Анализ представленной на рис. 1 информации показывает рост количества DDoS-атак относительно предыдущего отчетного периода. В свою очередь, по рекомендациям, данным в отчетах компании Positive Technologies, важно не только выстроить регулярные процессы определения и устранения уязвимостей, но и знать о существовании новых атак, следовательно, уметь быстро на них реагировать [3].

Для противодействия атакам, направленным на отказ в обслуживании, в общем случае необходимо идентифицировать тип трафика, который загружает сеть, а затем разделить поток на вредоносный и обычный [4].

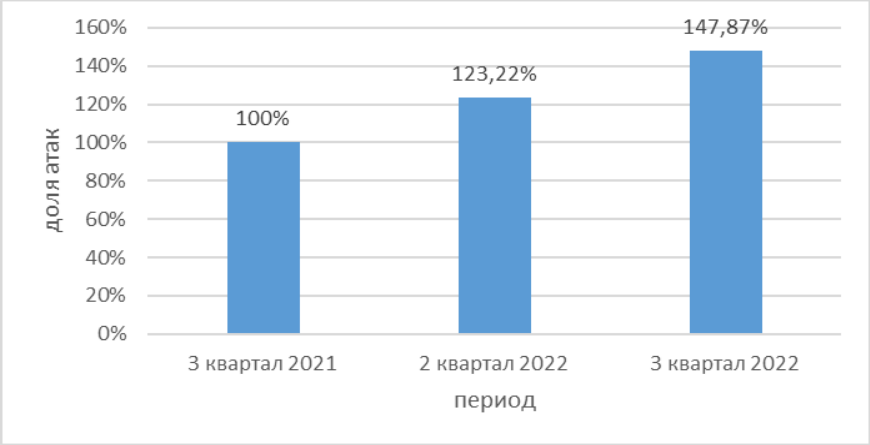

Рис. 1. Доля DDoS-атак в сравнении

В данной работе на примере статистических данных был проведен обзор методов прогнозирования аномалий.

2. Возможность использования энтропии для обнаружения DDoS-атак

Одним из признаков DDoS-атаки можно считать нетипично замедленную работу сетевого оборудования, что заметно без применения дополнительного анализа.

В [5] предложен алгоритм обнаружения аномалий, основанный на энтропии. Авторы показывают, что разработанный алгоритм обладает низкими вычислительными затратами, легкий в реализации, при этом обладает высокой скоростью обнаружения сетевых аномалий в режиме реального времени. Суть метода заключается в анализе влияния атаки на энтропию IP-адресов. DDoS-атака представляет собой большое количество запросов к конкретному сервису от одного узла-источника, то есть в общем трафике можно увидеть большое количество пакетов с одинаковыми IP-адресами – источника атаки и атакуемого сервера [6]. Атака за счет концентрации трафика на портах источника и портах назначения характеризуется уменьшением энтропии IP-адресов источника и IP-адресов назначения.

В качестве исходных данных для расчета энтропии воспользуемся базой KDD-2009 [7]. Из базы данных извлекаем записи, относящиеся к нормальному трафику и к DDoS-атакам. Далее из записи выделяем признак, относящийся к числу соединений с тем же самым IP-адресом порта назначения. На рис. 2 показано влияние DDoS-атаки на энтропию IP-адресов назначения.

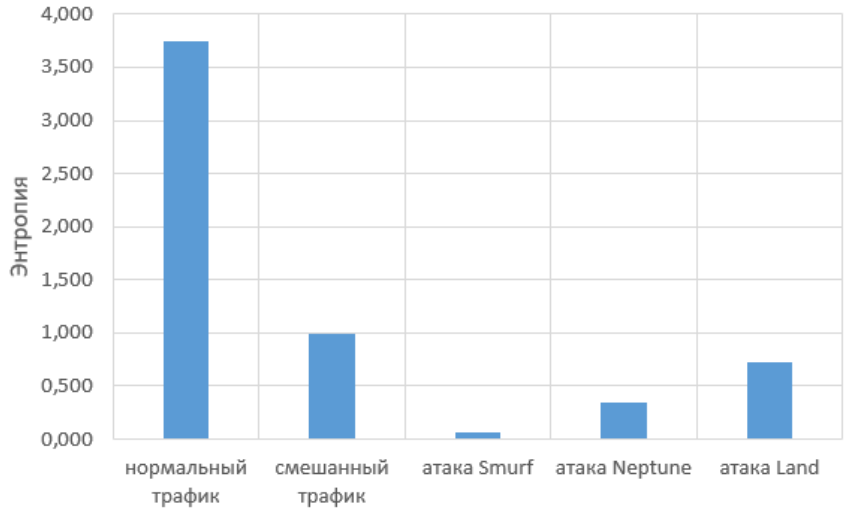

Рис. 2. Влияние DDoS-атаки на энтропию IP-адресов назначения

Анализ рис. 2 показал, что для нормального режима работы сети величина энтропии значительно превышает значения, рассчитанные для DDoS-атак. Таким образом, когда происходит DDoS-атака, число запросов на один IP-адрес резко увеличивается, что приводит к меньшему значению энтропии.

3. Прогнозирование DDoS-атак статистическими методами

Анализ исходных данных проводился на отрезке временного ряда данных с 1 ноября 2020 г. по 30 сентября 2022 г. [2].

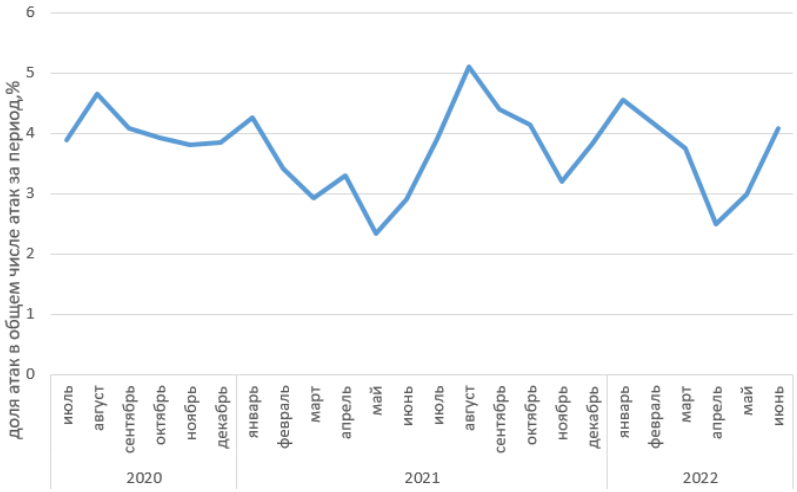

Рис. 3. Динамика DDoS-атак

Для прогнозирования динамики DDoS-атак используются различные методы и модели, которые различаются не только сложностью реализации, но и программной поддержкой. То есть выбор метода зависит от типа и цели прогноза, полноты исходных данных, доступности программного обеспечения и т.д.

В литературе описано большое количество методов статистического анализа и прогнозирования. К адаптивным моделям относят модель Брауна, модель Хольта и модель авторегрессии [8].

Метод экспоненциального сглаживания, на котором основаны модели Брауна и Хольта, позволяет анализировать временной ряд без предварительного задания уравнения тренда.

Основным моментом при использовании метода экспоненциального сглаживания является выбор параметра сглаживания α, начальных условий и степени полинома.

Если параметр сглаживания близок к нулю, значит, веса убывают медленно, и модель учитывает все значения рассматриваемого временного ряда. Напротив, если параметр близок к единице, это приведет к учету в прогнозе в основном влияния лишь последних наблюдений [8].

При разных значениях параметров сглаживания α результаты прогноза будут отличаться. Следовательно, параметр α выбирается таким образом, чтобы минимизировать ошибку прогноза.

Исследования показывают, что прогноз, учитывающий только один параметр α, нельзя считать абсолютно надежным. Для того, чтобы повысить точность прогноза, применяется модифицированная модель:

$$F_t(y) = \alpha \cdot y_t + (1-\alpha)(F_{t-1}(y) + T_{t-1}), \quad (1)$$

где выражение для тренда

$$T_t(y) = \beta \cdot (F_t - F_{t-1}) + (1-\beta) \cdot T_{t-1}, \quad (2)$$

где β – сглаживающая постоянная для тренда.

Меру отклонения прогноза от фактических значений можно оценить с помощью стандартной ошибки прогнозирования. В процессе настройки модели подбираются такие α и β, при которых стандартная ошибка отклонения будет минимальна. При этом необходимо учитывать условие применимости модели: если вышеописанные параметры принимают значения больше 0.7, прогноз нельзя считать достоверным.

Для расчета ошибки на основании исходных данных (рис. 3) подставляем в расчетную модель каждый параметр, изменяя его значение с 0 до 0.6 с шагом 0.1. Результаты расчета представлены на рис. 4.

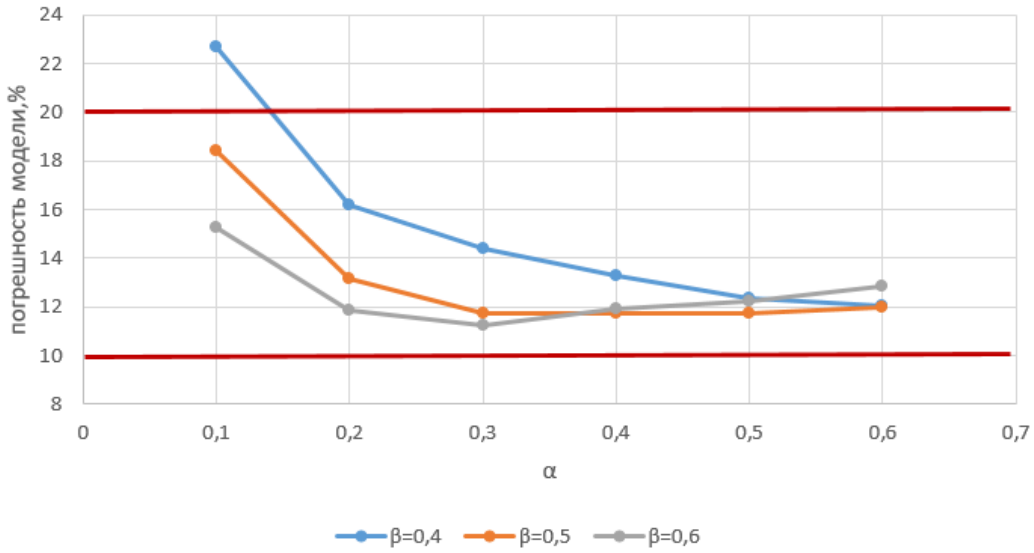

Рис. 4. Зависимость абсолютной ошибки модели от параметра сглаживания α для различных значений β

На рис. 4 видно, что минимальная ошибка достигается при α = 0.3, β = 0.6. Результаты расчета ошибки прогноза на тестовых данных приведены на рис. 5. Анализ рис. 5 показал, что минимальная ошибка тестирования достигается при α = 0.2, β = 0.4.

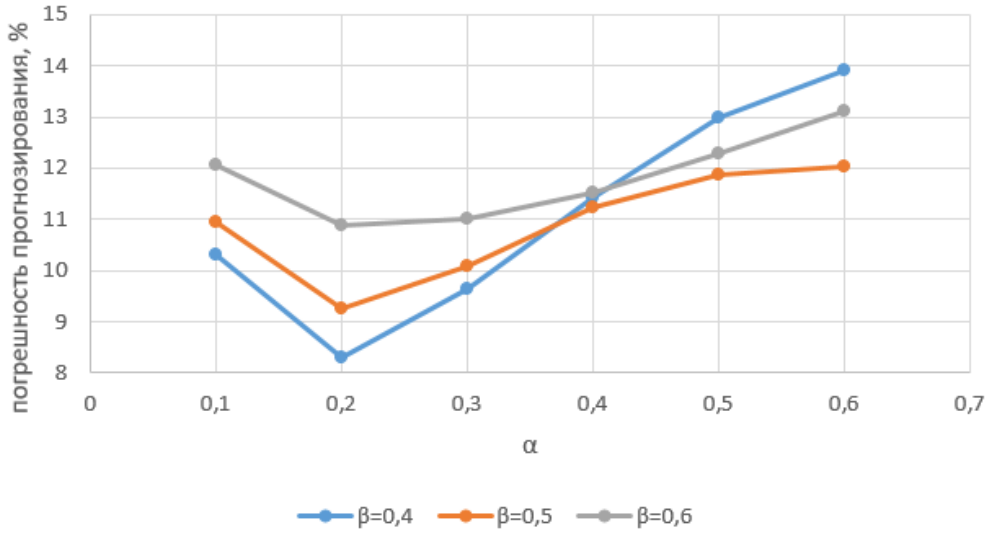

Рис. 5. Зависимость абсолютной ошибки тестирования модели от параметра сглаживания α для различных значений β

Так как значения α и β на обучающем и тестовом множествах не совпадают, учитывая, что для хорошей точности абсолютная ошибка находится в пределах от 0 % до 10 %, выбираем средние значения параметров, то есть $\alpha = 0.3$, $\beta = 0.5$.

В предложенных Дж. Боксом и Г. Дженкинсом моделях, в отличие от моделей, рассмотренных выше, применяется индивидуальный подход к каждому ряду. Существуют следующие модели Бокса–Дженкинса: авторегрессионная модель, модель скользящего среднего, смешанная модель с авторегрессией и скользящим средним, интегрированная модель авторегрессии (ARIMA) [5].

При построении ARIMA требуется анализ и выбор параметров модели. Кроме того, данные временных рядов должны быть стационарными, чтобы исключить корреляцию и мультиколлинеарность. Если исходный ряд не является стационарным, то его следует привести к стационарной форме.

Используя данные рис. 3, определим сезонный лаг при помощи спектрального анализа Фурье. Для этой цели был получен график периодограммы (рис. 6).

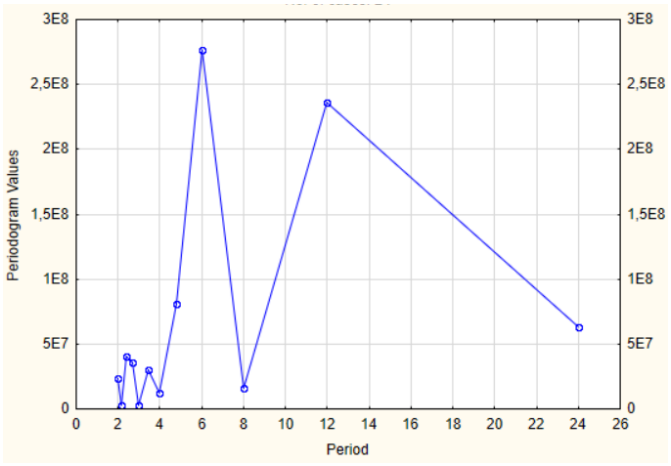

Рис. 6. График периодограммы

На графике видно максимальное значение в точке 6, то есть это значение является определяющим период сезонной составляющей рассматриваемого ряда.

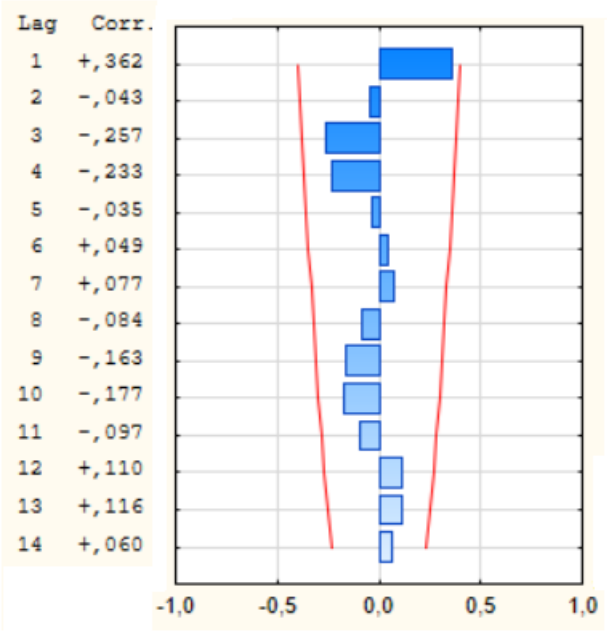

Рис. 7. Вычисленная коррелограмма преобразованного ряда

Так как коррелограмма ряда затухает с ростом лага и не превышает значения |0.3|, можно сделать вывод, что ряд обладает свойством стационарности, и для прогнозирования можно использовать модель ARIMA.

Расчеты показали, что ошибка моделирования составляет 9.1 %, что не превышает 10 %, позволяя выбрать эту модель для дальнейшей работы.

Таким образом, рассмотренные методы включают в себя следующие этапы построения: настройку моделей; отображение прогнозируемых данных в табличном или графическом виде; тестирование модели и расчет ошибки прогнозирования.

Литература

1. ГОСТ Р 53114-2008. Обеспечение информационной безопасности в организации [Электронный ресурс]. URL: https://docs.cntd.ru/document/1200075565 (дата обращения: 22.11.2022).
2. Лаборатория Касперского. Отчеты [Электронный ресурс] URL: https://www.kaspersky.ru/enterprise-security/resources (дата обращения: 22.11.2022).
3. Positive Technologies. Аналитика [Электронный ресурс] URL: https://www.ptsecurity.com/ruru/research/analytics/ (дата обращения 22.11.2022).
4. Методы защиты от DDOS нападений [Электронный ресурс]. URL: http://www.securitylab.ru/analytics/216251.php (дата обращения: 22.11.2022).
5. *Shanshan Yu, Jicheng Zhang, Ju Liu, Xiaoqing Zhang, Yafeng Li, Tianfeng Xu*. A cooperative DDoS attack detection scheme based on entropy and ensemble learning in SDN [Электронный ресурс]. URL: https://www.researchgate.net/publication/348891807 (дата обращения: 22.11.2022).
6. *Jung Woo Seo, Sangjin Lee*. A study on efficient detection of_network-based IP spoofing DDoS and malware-infected Systems [Электронный ресурс]. URL: https://www.researchgate.net/publication/309467794 (дата обращения: 22.11.2022)
7. The NSL-KDD Data Set. [Электронный ресурс]. URL: https://www.unb.ca/cic/datasets/nsl.html (дата обращения 22.11.2022).
8. *Афанасьев В. Н*. Анализ временных рядов и прогнозирование: учебник. Саратов: Ай Пи Ар Медиа, Оренбург: Оренбургский гос. ун-т, 2020. 286 с.

Лизнев Денис Сергеевич

аспирант СибГУТИ (630102, Новосибирск, ул. Кирова, 86), e-mail: liznev.denis@gmail.com, ORCID ID**:** 0009-0003-2599-8989.

Автор прочитал и одобрил окончательный вариант рукописи.
Автор заявляет об отсутствии конфликта интересов.

Overview of the Methods for Predicting Network Anomalies

Denis S. Liznev

Siberian State University of Telecommunications and Information Science (SibSUTIS)

Abstract: In this paper, the methods of predicting network anomalies are analyzed. Using the example of real statistical data, the stages of setting up forecasting models are shown. The effect of a DDoS attack on the destination IP-addresses' entropy is shown.

Keywords: exponential smoothing model, autoregressive model, entropy, network attacks.

For citation: Liznev D. S. Overview of the methods for predicting network anomalies (in Russian). *The SibSUTIS Bulletin*, 2023, vol. 17, no. 2, pp. 44-50. https://doi.org/10.55648/1998-6920-2023-17-2-44-50.

The article was submitted: 26.12.2022;
accepted for publication 10.01.2023.

References

1. GOST R 53114-2008. *Obespechenie informacionnoj bezopasnosti v organizacii* [Information security provision in organization], available at: https://docs.cntd.ru/document/1200075565 (accessed 22.11.2022).
2. *Laboratoriya Kasperskogo. Otchety* [DDoS reports], available at: https://www.kaspersky.ru/enterprise-security/resources (accessed 22.11.2022).
3. *Positive Technologies. Analitika* [Analytics], available at: https://www.ptsecurity.com/ruru/research/analytics/ (accessed 22.11.2022)
4. *Metody zashchity ot DDOS napadenij* [Methods of protection against DDOS attacks], available at: http://www.securitylab.ru/analytics/216251.php (accessed 22.11.2022)
5. Shanshan Yu, Jicheng Zhang, Ju Liu, Xiaoqing Zhang, Yafeng Li, Tianfeng Xu. *A cooperative DDoS attack detection scheme based on entropy and ensemble learning in SDN*, available at: https://www.researchgate.net/publication/348891807 (accessed 22.11.2022)
6. Jung Woo Seo, Sangjin Lee. *A study on efficient detection of_network-based IP spoofing DDoS and malware-infected Systems*, available at: https://www.researchgate.net/publication/309467794 (accessed 22.11.2022)
7. *The NSL-KDD Data Set*, available at: https://www.unb.ca/cic/datasets/nsl.html (accessed 22.11.2022)
8. Afanas'ev V. N. *Analiz vremennyh ryadov i prognozirovanie* [Time series analysis and forecasting]: Saratov, Aj Pi Ar Media, Orenburg, Orenburgskij gos. un-t, 2020. 286 p.

Denis S. Liznev
Postgraduate student, Siberian State University of Telecommunications and Information Science (SibSUTIS, Novosibirsk, Russia), e-mail: liznev.denis@gmail.com, ORCID ID: 0009-0003-2599-8989.