Научная статья
УДК 004.056; 519.25
https://doi.org/10.31854/1813-324X-2025-11-1-99-112
EDN:OOPJJR



# Анализ и прогнозирование временных рядов кибератак на информационную систему ведомственного вуза: возможности и ограничения методов

**Владимир Николаевич Наумов**[1], naumov122@list.ru
**Михаил Викторович Буйневич**[2] ✉, bmv1958@yandex.ru
**Максим Юрьевич Синещук**[2], smaxim@igps.ru
**Марина Алексеевна Тукмачева**[2], mtukmacheva@mail.ru

[1]Северо-Западный институт управления – филиал РАНХиГС,
Санкт-Петербург, 199178, Российская Федерация
[2]Санкт-Петербургский университет ГПС МЧС России,
Санкт-Петербург, 196105, Российская Федерация

**Аннотация**

***Актуальность статьи*** *обусловлена ростом угроз компьютерной безопасности критических информационных ресурсов, в том числе в системе образования, разнообразием видов и направлений кибератак, требующих дифференциации известных методов анализа и прогнозирования, в том числе на основе использования теории временных рядов.* ***Целью*** *статьи является исследование возможностей и ограничений использования методов теории временных рядов для анализа и прогнозирования динамики кибератак на примере ведомственного вуза, готовящего специалистов многим видам безопасности: техносферной, пожарной, информационной и проч. Высказана и проверена гипотеза о влиянии характера исходных данных на выбор методов анализа и прогнозирования временных рядов числа кибератак, о первичности исходных данных на результативность решения указанных задач. Выполнен анализ логов мониторинга межсетевого экрана корпоративной информационной системы; на их основе построены временные ряды числа различных видов атак и решены задачи текущего прогнозирования.* ***Новизна*** *полученных результатов обусловлена применением известных методов теории прогнозирования временных рядов к задаче исследования динамики кибератак на корпоративную информационную систему ведомственного вуза.* ***Теоретическая значимость*** *состоит в установлении границ возможности их применения в силу вариативности исследуемых временных рядов, а также в подтверждении первичности качества исходных данных над существующими методами и моделями.* ***Практическая ценность*** *определяется построением моделей временных рядов, позволяющих решать задачи текущего прогнозирования числа кибератак.*

**Ключевые слова:** *кибератаки, ведомственная информационная система, логи программно-аппаратного межсетевого экрана, временные ряды, анализ и прогнозирование, стационарность временных рядов, фильтры экспоненциального сглаживания, модели авторегрессии проинтегрированного скользящего среднего, метод Prophet*

**Источник финансирования:** *Работа выполнена в рамках НИР «Кибермониторинг» рег. № НИОКТР 1024040800041-6-2.2.66.*

**Ссылка для цитирования:** Наумов В.Н., Буйневич М.В., Синещук М.Ю., Тукмачева М.А. Анализ и прогнозирование временных рядов кибератак на информационную систему ведомственного вуза: возможности и ограничения методов // Труды учебных заведений связи. 2025. Т. 11. № 1. С. 99–112. DOI:10.31854/1813-324X-2025-11-1-99-112. EDN:OOPJJR

Original research
https://doi.org/10.31854/1813-324X-2025-11-1-99-112
EDN:OOPJJR

# Analyzing and Predicting the Time Series of Cyberattacks on Higher Education Departmental Institution Information System: Methods Opportunities and Limitations

**Vladimir N. Naumov**[1], naumov122@list.ru
**Mikhail V. Buinevich**[2] ✉, bmv1958@yandex.ru
**Maksim Y. Sineshchuk**[2], smaxim@igps.ru
**Marina A. Tukmacheva**[2], mtukmacheva@mail.ru

[1]North-West Institute of Management of the Russian Presidential Academy of National Economy and Public Administration,
St. Petersburg, 199178, Russian Federation
[2]Saint Petersburg University of State Fire Service of Emercom of Russia,
St. Petersburg, 196105, Russian Federation

**Annotation**

***The article relevance*** *is due to the growing threats to computer security of critical information resources, including in the education system, cyberattacks types and trends diversity, requiring known analysis and forecasting methods differentiation, including those based on the use of time series theory.* ***The article aim*** *is to study the possibilities and limitations of using time series theory methods to analyses and predict the cyber attacks dynamics on the departmental university example that trains specialists in many security types: technosphere, fire, information and other. Hypothesis about the influence of the initial data nature on the methods for cyberattacks number time series analyzing and forecasting choice, and primacy of initial data on the solving these tasks effectiveness was stated and tested. Analyses of the corporate information system firewall monitoring logs are performed. On their basis, time series number of different types of attacks are constructed. The tasks of building mathematical models and current forecasting have been solved. An integrated approach to their solution based on preliminary processing, testing of statistical hypotheses about DS- and TS-stationarity and use of different forecasting methods was applied. The obtained* ***results novelty*** *is due to known methods of time series forecasting theory application to studying the dynamics of cyberattacks on the departmental university corporate information system.* ***Theoretical significance*** *consists in establishing the limits of their application possibility due to the studied time series variability, as well as in confirming the initial data primary quality over the existing methods and models. The* ***practical value*** *is determined by the time series models construction that allow solving tasks of cyberattacks number current forecasting.*

**Keywords:** *cyberattacks, departmental information system, firewall logs, time series, analysis and forecasting, stationarity of time series, exponential smoothing filters, auto-regression models of the pro-integrated moving average, Prophet method*

**Funding:** *The work was carried out under the R&D "Cybermonitoring" Reg. No. NIOCTR 1024040800041-6-2.2.66.*

**For citation:** Naumov V.N., Buinevich M.V., Sineshchuk M.Y., Tukmacheva M.A. Analyzing and Predicting the Time Series of Cyberattacks on Higher Education Departmental Institution Information System: Methods Opportunities and Limitations. *Proceedings of Telecommunication Universities.* 2025;11(1):99–112. (in Russ.) DOI:10.31854/1813-324X-2025-11-1-99-112. EDN:OOPJJR

## Введение

Патриарх экономико-математической теории О. Моргерштерн указывал, что в триаде основных типов задач, решаемых учеными-исследователями, а именно – анализе, моделировании и прогнозировании, последняя является наиболее сильным вариантом постановки исследовательской проблемы [1]. Существует большое количество методов прогнозирования, качество которых зависит от имеющихся исходных данных. Если данные о прошлом представлены в числовой форме, а также имеются некоторые предположения, что выявленная на основе исследования ретроспективных данных тенденция может быть пролонгирована, то в этом случае используются количественные методы и, в частности, методы теории временных рядов.

За последние годы в ней разработано большое количество методов, алгоритмов и моделей, издается международный журнал прогнозирования, опубликовано множество книг, например [2, 3]. Создано значительное число пакетов прогнозирования для языков Python, R, что позволяет автоматизировать решение задач прогнозирования. Существующие статистические пакеты и графические надстройки, например, gretl, Loginom, JASP, jamovi позволяют в ходе исследования применять low-code, no-code подходы.

Популярность количественных методов прогнозирования увеличивается с возрастанием количества наборов данных, их размера (десятки гигабайт, например Kaggle). Только для решения задач прогнозирования на момент написания статьи их число превысило 8000. Более 500 датасетов посвящено проблемам кибербезопасности. Набор данных машинного обучения Калифорнийского университета UC Irvine Machine Learning Repository содержит 90 временных рядов, позволяющих решать задачи прогнозирования.

Анализ данных различной природы, включая и временные ряды, основан на модели, предложенной Дж. Тьюки, который утверждал, что «не метод определяет схему исследования, а характер данных». Вместо традиционно используемой последовательности исследования «модель – анализ – данные – результат», им была предложена схема «данные – разведочный анализ – модель – подтверждающий анализ». Первичной в этой схеме являются данные: их характер, используемые шкалы, объем, учет времени, качество, – все это определяет выбор инструмента исследования. Поэтому большинство публикаций, посвященных решению задач прогнозирования, в том числе и кибератак, непосредственно связано с характером исследуемых данных, наличием в них тренда, сезонных составляющих, характера случайной компоненты и др.

Так, например, в статье [4] исследуется динамика кибератак на веб-сервисы корпоративной сети, в том числе профили атак для различных стран. В основном использованы методы и инструменты графического и корреляционного анализа при допущении о стационарности исследуемых временных рядов. В [5] основное внимание уделяется прогнозированию общего числа атак, а также атак из определенных географических регионов на «сеть-приманку» (honeynet). Авторы использовали несколько подходов, таких как экспоненциальное сглаживание, ARIMA, SARIMA, GARCH и Bootstrapping. При этом показано, что различные методы обеспечивают различную точность для разных временных рядов. В [6] решена задача моделирования сезонных временных рядов количества атак на прикладное программное обеспечение с помощью гармонических составляющих. В этих статьях была дана характеристика источников данных и методики их предобработки.

В настоящей статье выполнен анализ динамики кибератак на информационную систему ведомственного вуза. Для получения исходных данных была использована BI-платформа, позволяющая графически представить динамику временных рядов в разрезе различных видов из разных стран. Периодичность поступления данных от источников позволяет сформировать временные ряды, получить многомерный временной ряд, сформировать панельные данные – то есть дополнительно к задачам визуального анализа решать задачи прогнозирования.

## Методы и инструменты

В качестве источников статистических данных инцидентов информационной безопасности были использованы логи программно-аппаратного межсетевого экрана IDECO UTM производства компании ООО «Айдеко». В зависимости от версий, этот межсетевой экран имеет спектр модулей фильтрации трафика: контент-фильтр, контроль приложений, предотвращение вторжений, антивирус веб-трафика. Рассматривались ряды количества атак по уровню угрозы. Столбиковые диаграммы числа кибератак, построенные с помощью ведомственной BI, приведены на рисунке 1. Другие ее визуальные элементы позволяют построить и исследовать временные ряды для различных стран и типов заблокированных атак (рисунок 2).

В ходе анализа рассматривались логи за 3 месяца межсетевого экрана. Полученные статистические данные позволили построить следующие интервальные временные ряды числа критических и опасных атак, числа предупреждений, а также – суммарного числа атак.

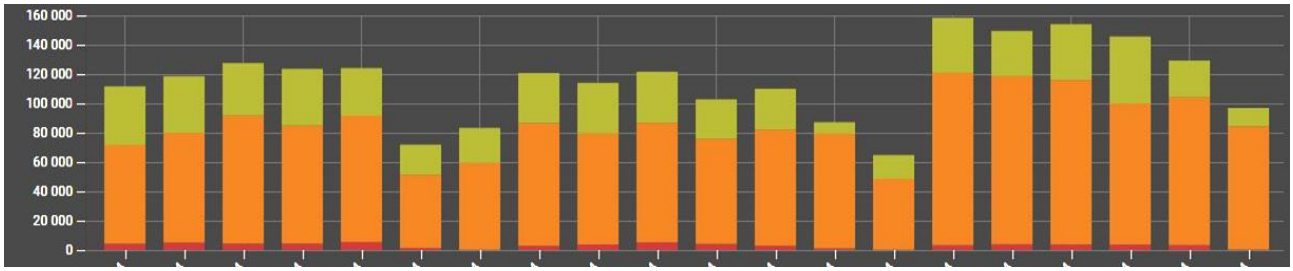
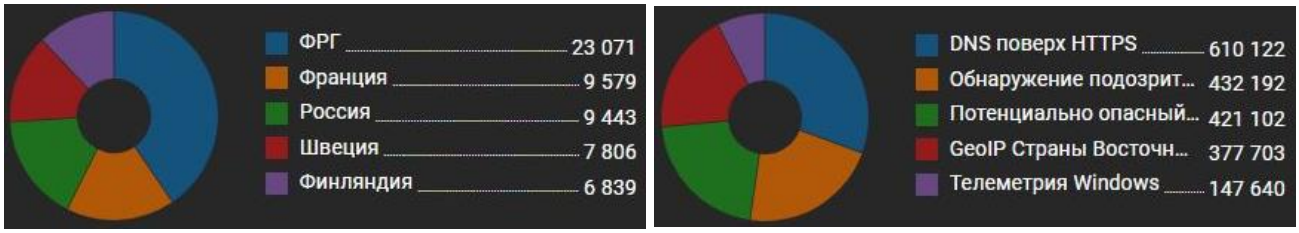



**Рис. 1. Количество атак по уровню угрозы**

*Fig. 1. Number of Attacks by Threat Level*




a) b)

**Рис. 2. Статистика кибератак на информационную систему ведомственного вуза: a) Топ атакующих стран; b) Топ заблокированных типов атак**

*Fig. 2. Statistics of Cyber Attacks on the Departmental University Departmental University Corporate Information System: a) Top Attacking Countries; b) Top Blocked Types of Attacks*

В качестве временного шага был выбран один час, что позволило построить сравнительно длинные ряды и сформулировать гипотезы о наличии сезонных составляющих, а также решить традиционные задачи разведывательного анализа: исследования стационарности временных рядов, построения их моделей и прогнозирования уровней исследуемых временных рядов с их помощью. В ходе исследования были использованы статистические пакеты JASP, jamovi, а также язык R и интегрированная среда разработки Rstudio. Выбор данных средств был обусловлен возможностью с их помощью автоматизировать большое количество задач прогнозирования, а в ряде случаев отказаться от разработки программных модулей. Реализовать технологию no-code. В частности, были применены следующие методы теории временных рядов:

– регрессионный анализ;

– экспоненциальное сглаживание;

– методы авторегрессии проинтегрированного скользящего среднего;

– байесовские методы пространства состояний;

– метод Prophet.

Такое большое количество методов позволило произвести сравнительный анализ результатов исследования, выбрать лучшие модели временных рядов и решить задачи прогнозирования с их помощью, а также дать характеристику динамики кибератак на исследуемую информационную систему. Так как данные методы разработаны на основе различных подходов, это позволяет учесть особенности анализируемых данных, реализовать модель Дж. Тьюки.

Результаты проведенного графического анализа исследуемых временных рядов с помощью Rstudio приведены на рисунке 3. Выполненный анализ показал, что временные ряды имеют сезонные составляющие, выявлена автокорреляция их уровней. Также установлено, что имеется большая дисперсия случайных составляющих, что усложняет их исследование.

## Результаты

Результаты описательной статистики данных временных рядов приведены в таблице 1. Диаграммы, представленные на рисунке 3, а также результаты описательной статистики позволяют сделать следующие выводы.

Во-первых, имеется большая вариативность уровней временных рядов. Для дальнейшего их исследования целесообразно решать задачи сглаживания или удалять аномальные наблюдения.

Во-вторых, коэффициент автокорреляции для каждого временного ряда значимо отличается от нуля. Следовательно, можно решать задачи прогнозирования.

В-третьих, коррелограммы показывают, что существуют сезонные составляющие временных рядов с периодом сезонности, равным 24 часам. Наибольшее число атак приходится на период с 9 до 15 часов, т. е. на дневное время. На рисунке 4 показаны «ящичные» диаграммы, которые были построены для исследуемых временных рядов.

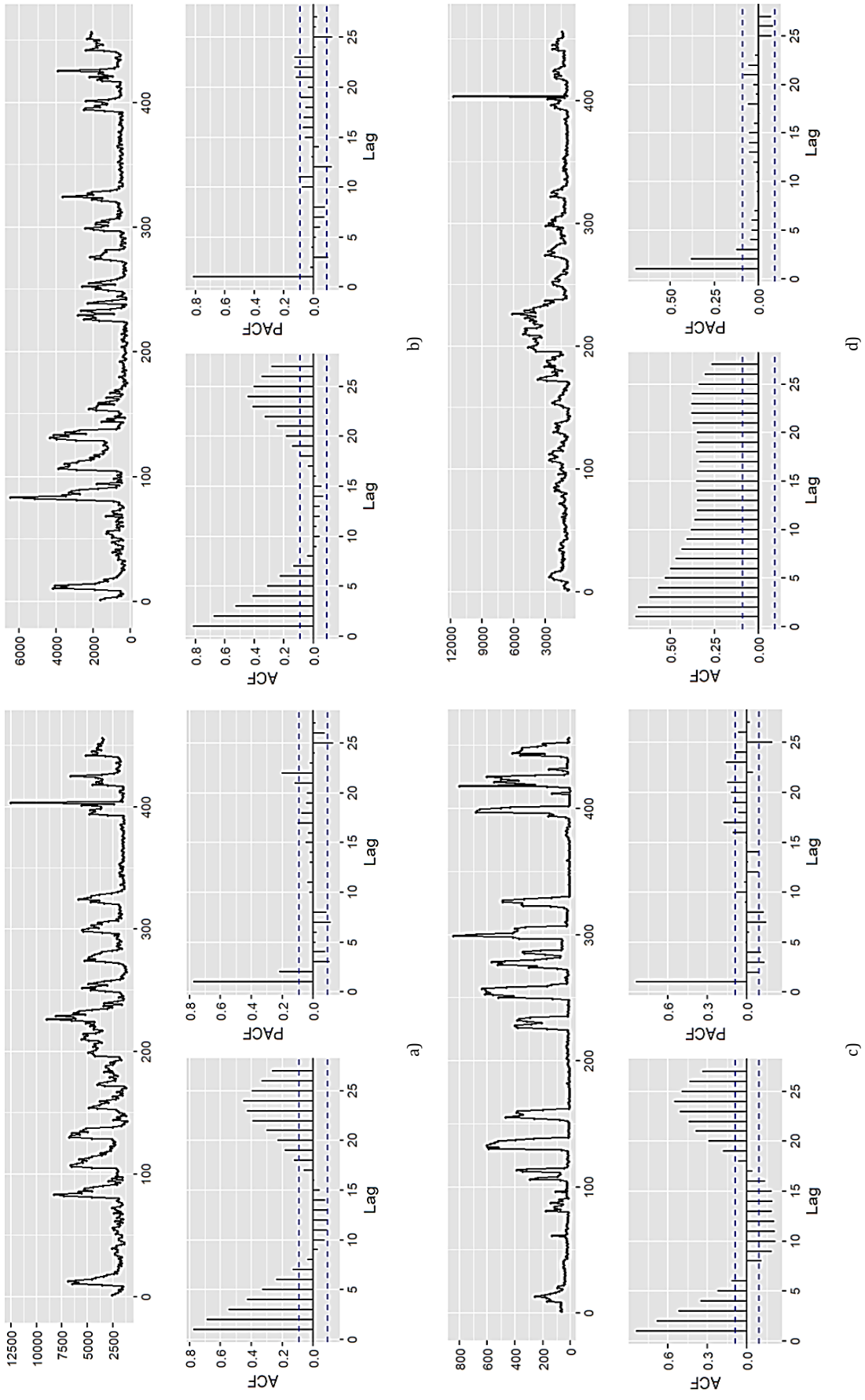



**Рис. 3. Точечная диаграмма и коррелограммы автокорреляционной (ACF) и частной автокорреляционной (PACF) функций: суммарного числа атак (a), числа предупреждений (b), критических (c) и опасных (d) атак**

*Fig. 3. Dot Plot and Correlograms of Autocorrelation Function (ACF) and Private Autocorrelation Function (PACF): Total Number of Attacks (a), Number of Warnings (b), Critical (c) and Dangerous (d) Attacks*

**ТАБЛИЦА 1. Описательная статистика**

*TABLE 1. Descriptive Statistics*

| Показатели | Временные ряды | | | |
|---|---|---|---|---|
| | Критичные | Опасные | Предупреждение | ВСЕГО |
| Существующие | 456 | 456 | 456 | 456 |
| Пропущенные | 0 | 0 | 0 | 0 |
| Среднее | 110,746 | 1705,908 | 1046,156 | 2862,809 |
| Стандартное отклонение | 169,834 | 1032,702 | 918,354 | 1566,028 |
| Дисперсия | 28 843,557 | 1 066 000 | 843 374,923 | 2 452 000 |
| Размах | 839 | 11013 | 6277 | 11441 |
| Минимум | 3 | 664 | 178 | 1060 |
| Максимум | 842 | 11677 | 6455 | 12501 |
| Start | 63 | 898 | 1597 | 2558 |
| End | 10 | 1414 | 2105 | 3529 |
| Автокорреляция первого порядка | 0,844 | 0,698 | 0,82 | 0,776 |

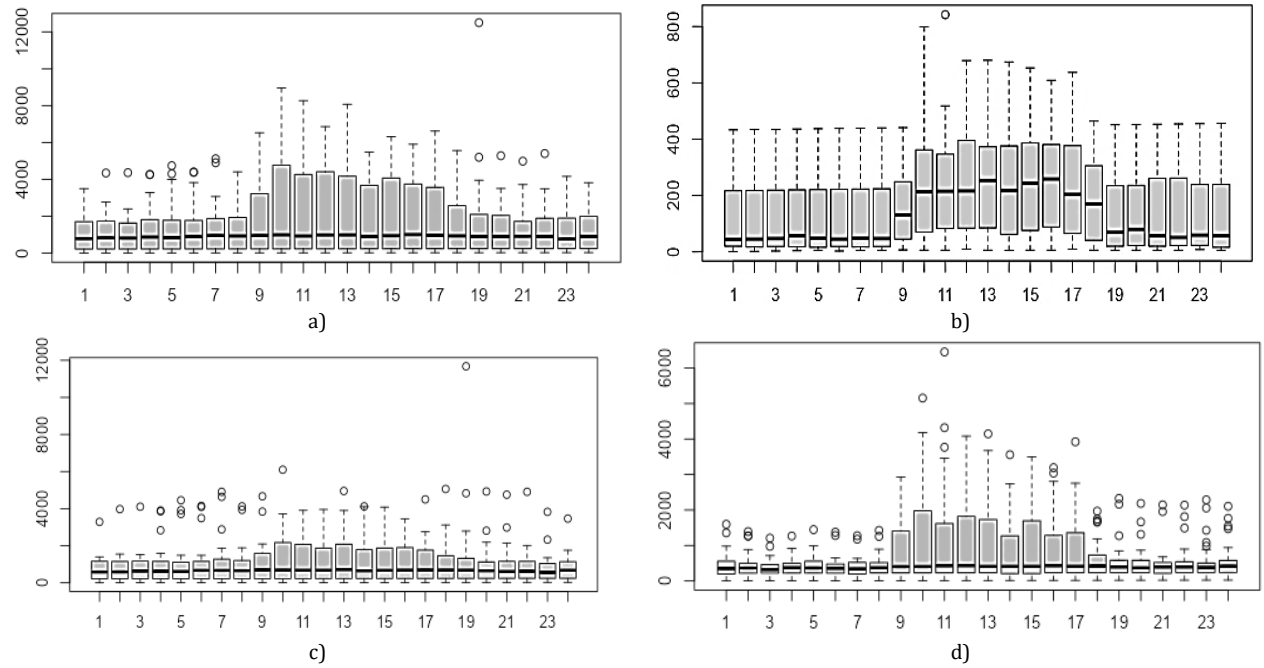



**Рис. 4. «Ящичные» диаграммы числа: a) всех атак; b) критических атак; c) опасных атак; d) предупреждений**

*Fig. 4. "Box" Plots of the Number of all Attacks (a); Critical Attacks (b); Dangerous Attacks (c) and Alerts (d)*

Наибольшие размеры «ящиков» (см. рисунок 4) также приходятся на дневное время. При этом характерно, что размах значений и длины верхних «усов» для анализируемого периода также максимальны. Диаграммы вновь подтверждают наличие выбросов, которые на диаграммах обозначены круглыми маркерами, расположенными над верхними «усами». Их число для разных рядов составляет от 19 до 33. Существуют и экстремальные значения, приходящиеся на 11 и 19 часов.

Для построения моделей временных рядов потребовалось выполнить анализ их стационарности с помощью статистических критериев Дикки – Фуллера, KPSS [7] и Филиппа – Перона [8]. Результаты проверки данных статистических гипотез приведены в таблице 2.

Данная таблица учитывает два вида стационарности временных рядов – DS- и TS-стационарность (*аббр. от англ.* Difference Stationary, разностно-стационарный и Trend Stationary, стационарный относительно тренда). В первом случае ряд является $I(k)$-интегрированным, например, случайным блужданием $I(1)$. Приведение его к стационарному осуществляется с помощью нахождения разностного ряда $k$-го порядка, т. е. получения так называемого $I(0)$-стационарного процесса. При TS-стационарности, частным случаем которой является $I(0)$-ряд (уровнево-стационарный ряд), из наблюдаемых значений необходимо вычесть значения детерминированной функции, описывающей тренд.

**ТАБЛИЦА 2. Статистические критерии проверки стационарности временных рядов**

*TABLE 2. Statistical Criteria for Testing Stationarity of Time Series*

| Временной ряд | Критерий | Значение критерия | Значение лага | Уровень значимости (*p*-value) | Проверяемая гипотеза (H0): ряд … |
|---|---|---|---|---|---|
| Предупреждения | Критерий Дики – Фуллера | –2,826 | 4 | 0,234 | не стационарен |
| | Критерий Филлипса – Перона | –20,355 | 3 | 0,055 | не стационарен |
| | Критерий KPSS, уровневая стационарность | 0,182 | 4 | 0,100 | уровнево-стационарен |
| | Критерий KPSS, стационарность тренда | 0,135 | 4 | 0,071 | стационарен по тренду |
| Опасные атаки | Критерий Дики – Фуллера | –3,479 | 4 | 0,047 | не стационарен |
| | Критерий Филлипса – Перона | –21,399 | 3 | 0,044 | не стационарен |
| | Критерий KPSS, уровневая стационарность | 0,160 | 4 | 0,100 | уровнево-стационарен |
| | Критерий KPSS, стационарность тренда | 0,147 | 4 | 0,050 | стационарен по тренду |
| Критические атаки | Критерий Дики – Фуллера | –3,092 | 4 | 0,124 | не стационарен |
| | Критерий Филлипса – Перона | –32,156 | 3 | 0,010 | не стационарен |
| | Критерий KPSS, уровневая стационарность | 0,517 | 4 | 0,038 | уровнево-стационарен |
| | Критерий KPSS, стационарность тренда | 0,206 | 4 | 0,014 | стационарен по тренду |
| Все атаки | Критерий Дики – Фуллера | –2,948 | 4 | 0,184 | не стационарен |
| | Критерий Филлипса – Перона | –18,364 | 3 | 0,086 | не стационарен |
| | Критерий KPSS, уровневая стационарность | 0,158 | 4 | 0,100 | уровнево-стационарен |
| | Критерий KPSS, стационарность тренда | 0,143 | 4 | 0,055 | стационарен по тренду |

Приведенное выше разнообразие видов стационарности определяет не только разнообразие применяемых статистических критериев, но и проверяемых статистических гипотез (последний столбец таблицы, содержащий описание нулевой статистической гипотезы). Отметим, что в четвертом столбце таблицы указано значение лага. Это позволяет при проверке стационарности использовать так называемые расширенные статистические тесты, предполагающие, что анализируемый случайный процесс не является авторегрессионным первого порядка, а описывается более сложной моделью с бóльшим, чем один числом лагов.

Значение уровней значимости (*p*-value) для критериев Дики – Фуллера и Филлипса – Перрона больше, например 0,05, позволяют сделать вывод, что временной ряд предупреждений является DS-стационарным. Чтобы его сделать уровнево-стационарным, необходимо построить ряд разностей. Итоговый временной ряд всех атак, а также временной ряд опасных атак не относятся к категории DS-рядов. С другой стороны, на уровне 0,055 они является уровнево-стационарными. Ряды не содержат тренда, и возможно, имеют ненулевое математическое ожидание своих уровней. И наконец, анализ стационарности временного ряда критических атак с помощью четырех критериев приводит к противоречиям: первые два критерия на уровне 0,05 не позволяют сделать вывод о DS-стационарности, а вторые два на этом же уровне значимости не отвечают на вопрос о стационарности по тренду или об уровневой стационарности. Напомним, что в этом случае возможно использовать поправку Бонферрони, являющуюся методом противодействия проблеме множественных сравнений при применении семейства статистических гипотез. Необходимо продолжить исследование, например, с помощью моделей ARIMA.

Сравнительный анализ результатов построения моделей для одного из анализируемых временных рядов (всех атак) разными методами приведен в таблице 3. Данные результаты показывают низкое качество для различных классов моделей. Здесь в качестве критериев оценки их качества использованы:

– показатель ранжированной оценки вероятности (CRPS, *аббр. от англ*. Continuous Ranked Probability Score) [9];

– показатель Дэвида – Себастьяни (DSS, *аббр. от англ*. Dawid-Sebastiani Score) [10];

– средняя абсолютная ошибка аппроксимации (MAE, *аббр. от англ*. Mean Absolute Error);

– квадратный корень из среднего квадрата ошибки аппроксимации (RMSE, *аббр. от англ*. Root Mean Squared Error);

– коэффициент детерминации ($R^2$).

**ТАБЛИЦА 3. Результаты оценки качества модели**

*TABLE 3. Results of Model Quality Assessment*

| Класс модели | CRPS | DSS | MAE | RMSE | $R^2$ |
|---|---|---|---|---|---|
| Линейная регрессия | 927,678 | 15,940 | 1405,956 | 1663,859 | 0,012 |
| Байесовская модель BSTS | 2901,963 | 19,174 | 1888,910 | 2454,455 | 0,108 |
| Байесовская авторегрессионная модель | 1235,123 | 16,621 | 1659,680 | 2127,404 | 0,115 |
| Prophet | 1426,080 | 19,452 | 1790,740 | 2154,794 | 0,381 |

Если последние три критерия применяются сравнительно часто, то первые два нуждаются в пояснении. Так, показатель CRPS – непрерывная ранжированная оценка вероятности, является обобщением показателя MAE для случая вероятностных прогнозов; его меньшему значению соответствует лучшая модель. Показатель DSS оценивает средние значения вектора отклонений наблюдаемых и прогнозных значений; здесь также меньшему значению критерия соответствует лучшая модель. Приведенные значения показывают, что нет лучшей по всем показателям модели, но по большинству показателей лучшей является линейная регрессионная модель, что довольно неожиданно. Однако ее построение и оценка качества такой модели также не позволяет сделать вывод о ее применимости.

Таким образом, без предварительной обработки с целью повышения качества исходных данных задача прогнозирования не может быть решена. Поэтому дальнейшее исследование было проведено с учетом необходимости повышения качества исходных данных за счет преобразования временных рядов. Известны различные методы таких преобразований, например, логарифмирование или извлечение квадратного корня наблюдаемых уровней. Их обобщением является преобразование Бокса – Кокса [11], при выполнении которого необходимо задать или найти значение параметра данного преобразования λ. Однако в этом случае затрудняется интерпретация полученных результатов, и возникает необходимость обратного преобразования.

В качестве альтернативы выберем методы фильтрации, в частности, метод ETS (*аббр. от англ.* Triple Exponential Smoothing, тройного экспоненциального сглаживания) [12]; система уравнений такого фильтра позволяет сгладить уровни временно́го ряда, тренд, а также сезонные составляющие. При этом модель задается трехзначным символьным кодом, первый знак которого определяет тип случайной составляющей «E», второй – тип тренда «T», третий – характеризует сезонную составляющую «S». Такой код позволяет задать пятнадцать классов фильтров сглаживания. Будем использовать средства подгонки лучшего фильтра и оптимизации значений его параметров; их число зависит от выбора класса фильтра.

Для построения моделей временных рядов выполним композицию двух методов: экспоненциального сглаживания и авторегрессии ARIMA. Возможности применяемых программных средств позволяют использовать методологию autoML и подобрать с ее помощью нужные значения гиперпараметров, как для фильтров, так и для моделей ARIMA. Так как ее параметры подбираются автоматически, например, по значению информационных критериев, то при их определении возможно получение частных видов модели, например ARMA, AR и MA. В случае необходимости следует использовать расширения – SARIMA, ARIMAX, SARIMAX: их возможности позволят исследовать влияние рядов критических, опасных и атак-предупреждений на их общее число.

Лучшая модель фильтра позволяет получить сглаженные значения уровней временного ряда $l_t$. По сглаженным значениям можно построить модель ARIMA, откликом для которой будет значение уровня ряда на момент времени $t$. Ее вид и значения гиперпараметров при этом определяются автоматически.

Первый временной ряд, содержащий все атаки, может быть представлен моделью ARIMA (1, 0, 2) с ненулевым математическим ожиданием:

$$l_t = 2858{,}52 + 0{,}82l_{t-1} + \varepsilon_t + 0{,}19\varepsilon_{t-2}, \varepsilon_t : N(0; 741).$$

Фильтр сглаживания относится к классу фильтра с мультипликативной ($M$) ошибкой («E» = $M$), с отсутствием ($N$ – None) тренда («T» = $N$) и сезонной составляющей («S» = $N$); его уравнение имеет вид:

$$l_t = 0{,}77y_t + 0{,}23l_{t-1}, l_{init} = 2246{,}2,$$

где $l_{init}$ – начальное состояние фильтра.

Построенная модель временного ряда позволяет выполнить интервальную оценку условного математического ожидания прогноза. На рисунке 5 приведена диаграмма прогнозирования на три дня с построением верхней и нижней границ 80- и 95-процентных доверительных интервалов. Сравнительно небольшая их ширина и медленный ее рост позволяют увеличить горизонт прогноза и с учетом знаков коэффициентов в уравнении предположить, что в среднем, в недалеком будущем общий поток атак не увеличится.

Аналогично решены задачи построения модели фильтра сглаживания, уравнения ARIMA и прогнозирования для других временны́х рядов. Так, временно́й ряд, содержащий критические атаки, может быть представлен моделью ARIMA (2, 0, 1) с ненулевым математическим ожиданием, имеющей вид:

$$l_t = 110{,}34 + 1{,}69l_{t-1} + 0{,}75l_{t-2} + \varepsilon_t - 0{,}77\varepsilon_{t-1},$$
$$\varepsilon_t : N(0{,}28; 8).$$

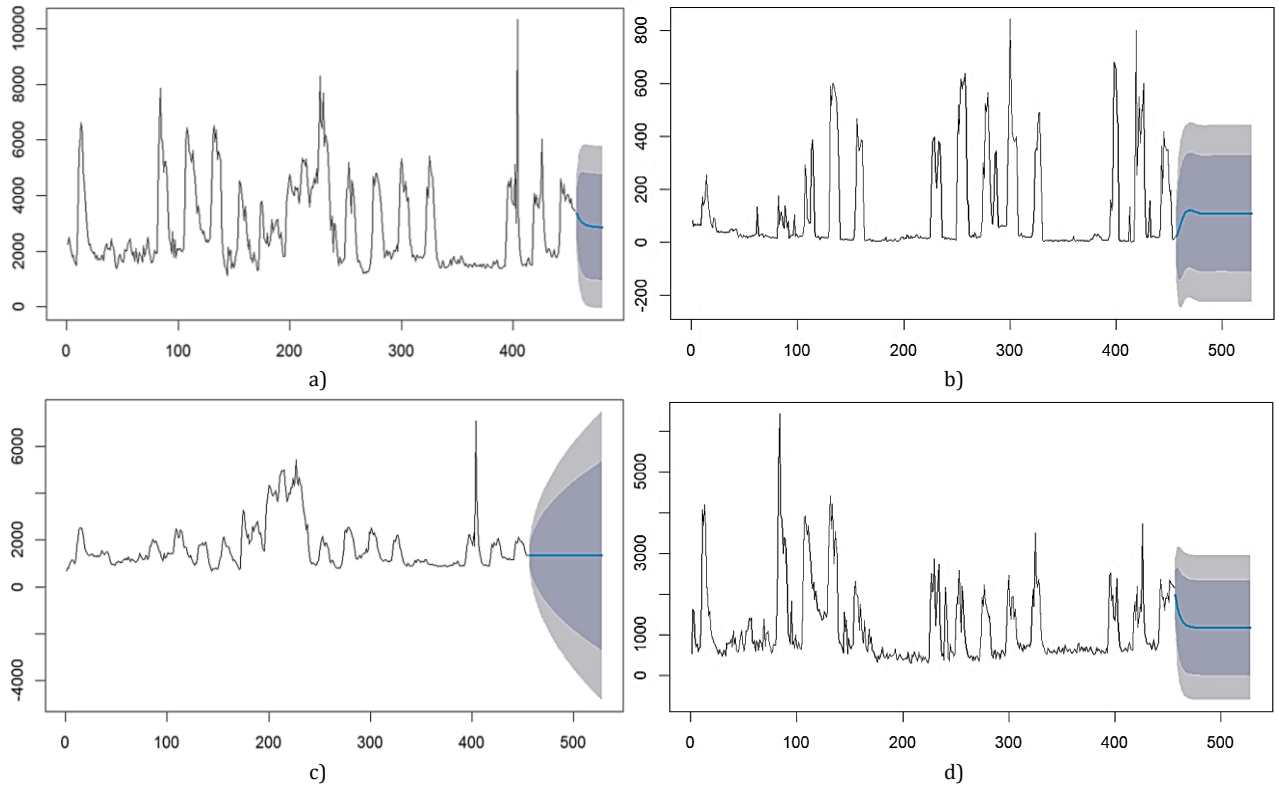



**Рис. 5. Диаграмма прогнозирования уровней временного ряда числа: a) всех атак; b) критических атак; c) опасных атак; d) предупреждений**

*Fig. 5. Time Series Level Prediction Diagram of the Number of All Attacks (a), Critical Attacks (b), Dangerous Attacks (c) and Alerts (d)*

Сглаженные уровни временного ряда $l_t$ были получены с помощью фильтра простого экспоненциального сглаживания с параметром α = 1 и инициальным значением фильтра, равным 83,0. К сожалению, данный фильтр не сглаживает уровни временного ряда, поэтому диаграмма, приведенная на рисунке 5b, показывает большую ширину обоих доверительных интервалов. Возможной причиной такой ситуации может быть большое число выбросов и большой разброс значений уровней временного ряда. Тем не менее, полученный на три дня кратковременный прогноз также позволяет сделать вывод о стационарности временного ряда, т. е. подтвердить результаты ранее проверенных статистических гипотез по различным критериям.

Временной ряд, содержащий сведения об опасных атаках, описывается моделью авторегрессии-скользящего среднего ARIMA (0, 1, 0):

$$l_t = l_{t-1} + \varepsilon_t;\ \varepsilon_t : N(0; 370).$$

Для сглаживания его уровней с помощью autoML был определен аддитивный фильтр тройного экспоненциального сглаживания ETS с мультипликативной случайной составляющей (MAN).

Уравнения данного фильтра сглаживания содержат два оцененных параметра для каждого из них и имеют вид:

$$l_t = 0{,}5l_{t-1} + 0{,}49(l_{t-1} + b_{t-1}),$$
$$b_t = 0{,}02(l_t - l_{t-1}) = 0{,}98b_{t-1},$$
$$l_{init} = 704{,}1, b_{init} = -3{,}53.$$

Инициальные значения уровня ряда $l_{init}$ и тренда $b_{init}$ позволяют применять данный фильтр для решения задач сглаживания.

Построенная модель ARIMA позволяет сделать вывод, что данный временной ряд является DS-рядом, т. е. имеет стохастический тренд. С ростом времени прогнозирования растет ширина доверительного интервала прогноза, что не позволяет решать задачи долгосрочного прогнозирования уровней временного ряда. Это подтверждается колоколообразным видом доверительных интервалов прогноза, приведенных на рисунке 5c.

Последний временной ряд также может быть представлен моделью ARIMA с параметром авторегрессии $p$ = 2 и параметром скользящего среднего $q$ = 2 с ненулевым математическим ожиданием, имеющей следующий вид:

$$l_t = 116827 + 0{,}69l_{t-1} + 0{,}01l_{t-2} + \varepsilon_t + 0{,}24\varepsilon_{t-1} + $$
$$+0{,}13\varepsilon_{t-2}, \varepsilon_t : N(0; 467).$$

Сглаженные значения данного временного ряда получены с помощью соотношений:

$$l_t = 0{,}88l_{t-1} + 0{,}12(l_{t-1} + b_{t-1}),$$

$$b_t = 0{,}001(l_t - l_{t-1}) + 0{,}999b_{t-1},$$
$$l_{init} = 376{,}1, b_{init} = 155{,}59.$$

Диаграмма прогнозирования уровней временного ряда на три дня (см. рисунок 5d) также показывает, что доверительный интервал прогноза сравнительно невелик, поэтому можно увеличивать горизонт прогноза.

Полученные модели прогнозирования сглаженных уровней позволяют сделать вывод, что все анализируемые временные ряды, кроме опасных атак, относящихся к нестационарному ряду «случайное блуждание», являются TS-стационарными, а их случайные составляющие могут быть описаны авторегрессионными зависимостями. Отметим, что все коэффициенты, приведенные в уравнениях моделей ARIMA, значимо отличаются от нуля на сравнительно высоком уровне. Данный вывод сделан с помощью статистического критерия Стьюдента.

Дальнейшее исследование может быть направлено на анализ компонентов временных рядов, в частности тренда, сезонной и случайной составляющей, несмотря на то, что фильтры экспоненциального сглаживания не позволили выявить сезонные составляющие. С этой целью целесообразно использовать метод Prophet, который основан на подгонке аддитивных регрессионных моделей, включающих тренд $g_t$, сезонные колебания $s_t$, эффекты праздников $h_t$, а также случайную составляющую $\varepsilon_t$. Его выбор основан на том, что он хорошо работает в условиях годовой, недельной и ежедневной сезонности, реализован в языках аналитики R, Python, а также в их графических приложениях.

В общем виде аддитивная регрессионная модель временного ряда, построенная с помощью данного метода, принимает вид:

$$y_t = g_t + s_t + h_t + \varepsilon_t,$$

где $g_t$ – тренд; $s_t$ – совокупность сезонных составляющих; $h_t$ – составляющая учитывающая эффекты праздников и других влиятельных событий; $\varepsilon_t$ – случайная компонента.

Для выявления сезонных составляющих $s_t$ используется разложение в ряд Фурье. При определении тренда применяются кусочные линейная или логистическая модели с использованием точек излома. Для построения модели и выявления ее компонент используем платформу jamovi, которую можно рассматривать как графическую надстройку языка R (https://www.jamovi.org). Результаты декомпозиции на основе построения кусочной линейной регрессионной модели с годовыми (yearly), недельными (weekly) и дневными (daily) колебаниями для исследуемых временных рядов приведены на рисунках 6–9.

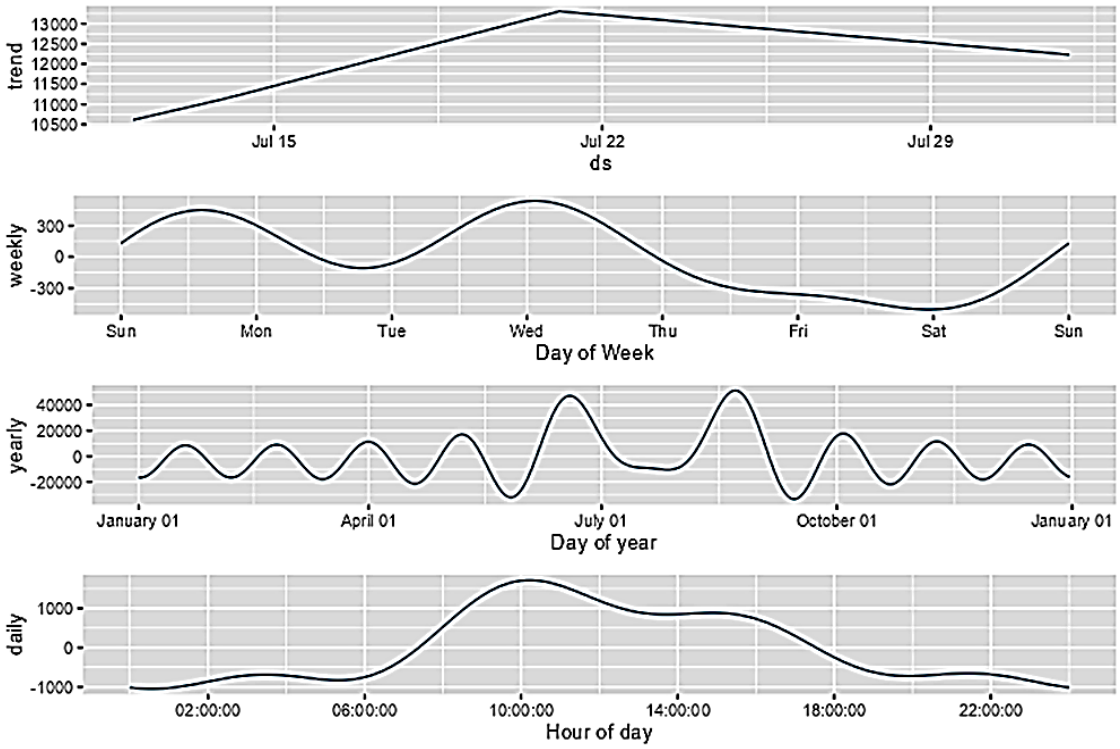



**Рис. 6. Декомпозиция временного ряда общего числа атак**

*Fig. 6. Time Series Decomposition of Total Attacks Number*

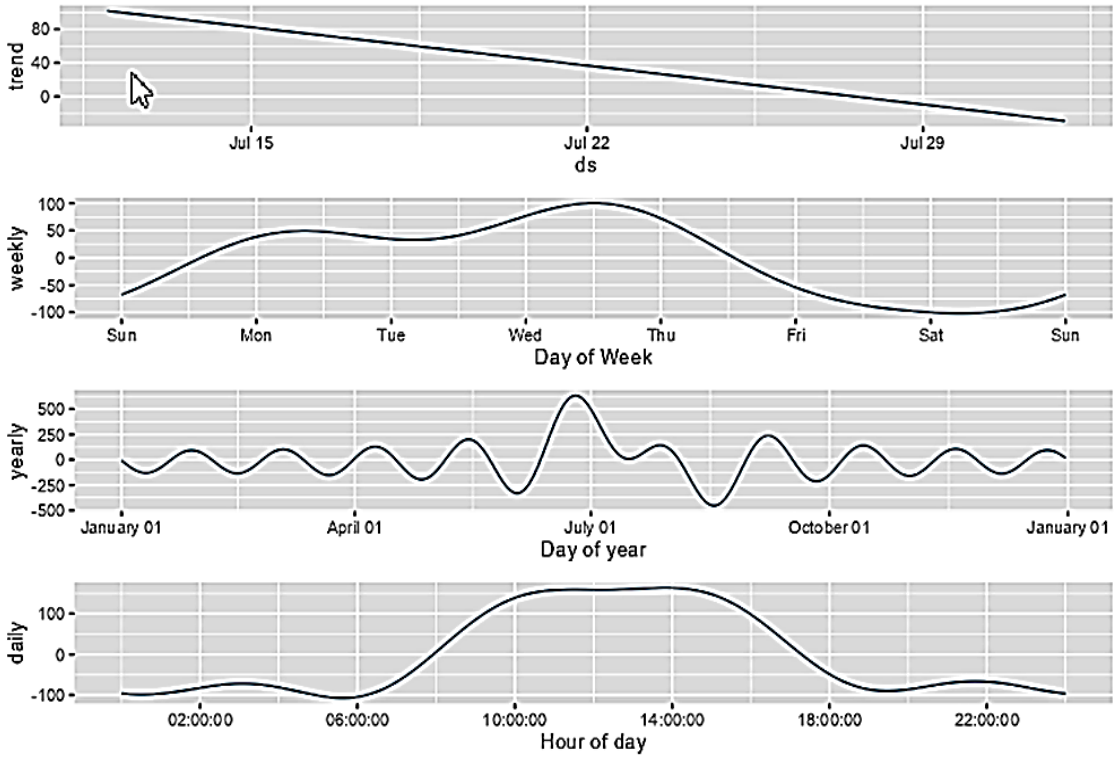



**Рис. 7. Декомпозиция временного ряда числа критических атак**

*Fig. 7. Time Series Decomposition of Critical Attacks Number*

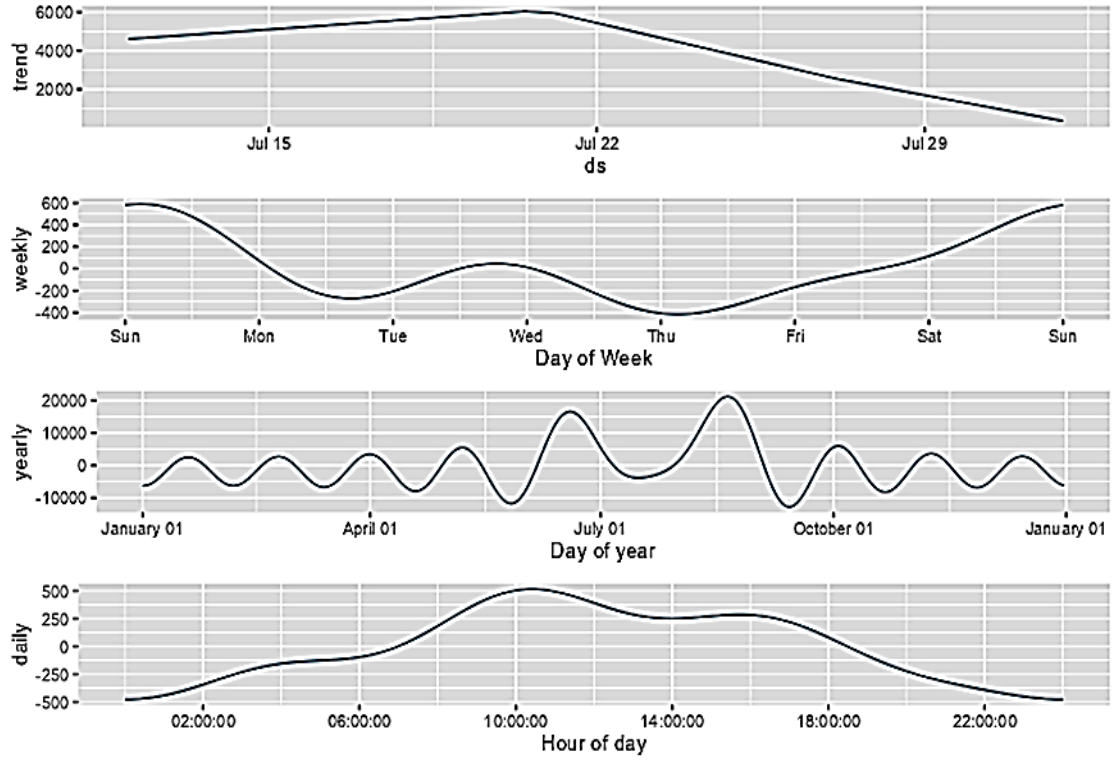
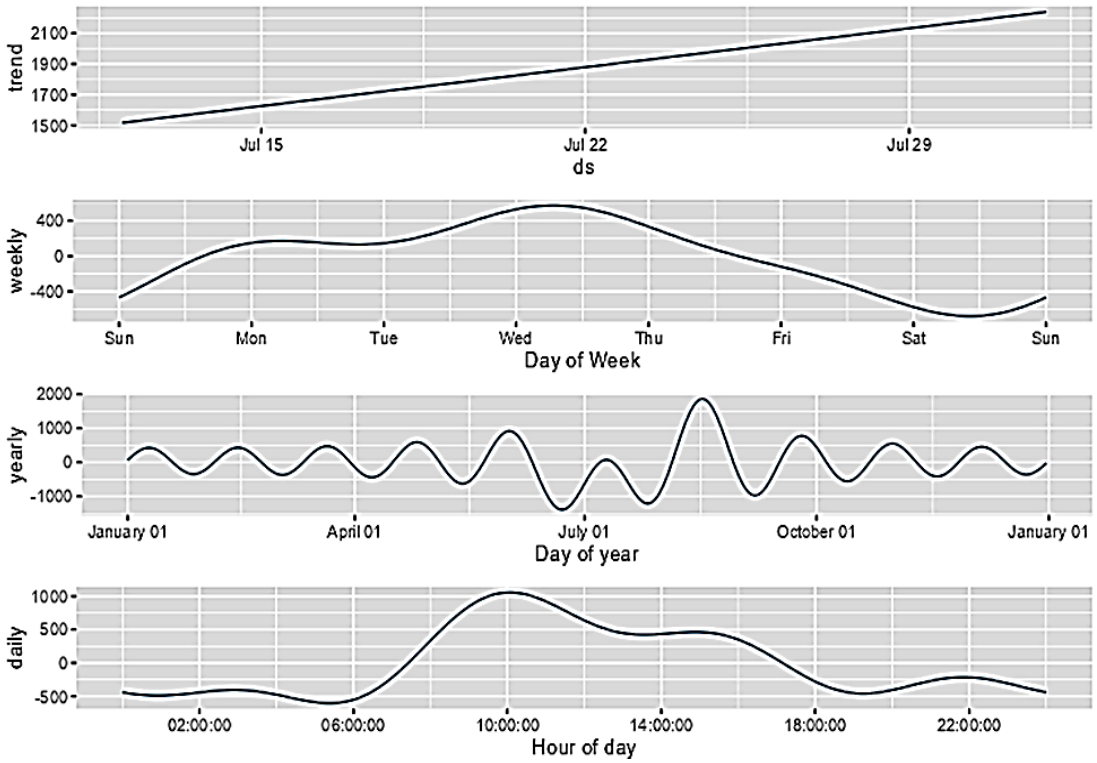



**Рис. 8. Декомпозиция временного ряда числа опасных атак**

*Fig. 8. Time Series Decomposition of Dangerous Attacks Number*




d)

**Рис. 9. Декомпозиция временного ряда числа предупреждений**

*Fig. 9. Time Series Decomposition of Alerts Number*

Декомпозиция показывает, что имеются точки излома тренда для двух временных рядов (общего числа атак и числа опасных атак). Таким образом, ряды могут содержать различные участки монотонности. Следовательно, с учетом их разведывательного анализа может возникнуть необходимость создавать их слайсы и для каждого слайса строить модель временного ряда. Вероятно, это сможет повысить качество модели.

Построенные модели сезонной декомпозиции показывают, что анализируемые ряды ведут себя по-разному. Так, например, для ряда с опасными атаками наибольшее число атак в среднем приходится на выходные. А для ряда, содержащего общее число атак и число критических атак – на среду. Эта информация является существенной для выбора и обоснования моментов повышенной готовности системы обнаружения и ликвидации последствий вторжений. Отметим, что для всех временных рядов, наибольшее число атак приходится на утренние часы, что также немаловажно для планирования работы системы обеспечения кибербезопасности.

## Заключение

Полученные результаты показывают, что при решении задач прогнозирования кибератак на объекты информационной инфраструктуры целесообразно использовать методы теории временных рядов. Ее разработанность, наличие большого количества методов и инструментальных средств позволяют реализовать комплексный подход, основанный на их последовательном применении, построении различных моделей, а затем их сравнительном анализе. Проведенное исследование четырех временных рядов на примере информационной системы ведомственного вуза показывает, что в силу большой вариативности, зашумленности измерений, наличия случайной составляющей с большой дисперсией использование традиционных подходов к прогнозированию (например, методов регрессионного анализа), как правило, не эффективно. Так, коэффициент детерминации построенной модели для одного из временных рядов составляет не более 0,01, а исправленное его значение становится даже отрицательным!

Попытка предобработать временные ряды с использованием преобразования Бокса – Кокса также не позволяет существенно улучшить качество решаемой задачи. Поэтому в исследовании использованы методы экспоненциального сглаживания, позволяющие уменьшить дисперсию временных рядов. В дальнейшем результаты фильтрации могут быть применены при решении задач прогнозирования.

К сожалению, даже это не позволяет существенно улучшить качество исходных данных. Поэтому решение задачи прогнозирования возможно

только для небольшого горизонта прогноза, который в исследовании был задан равным трем суткам. Вероятно, его можно увеличить, но предварительно следует исследовать характер анализируемого временного ряда и возможность такого подхода. В любом случае, большая зашумленность данных не позволяет решать задачи долгосрочного прогнозирования. Задача прогнозирования кибератак является задачей краткосрочного или текущего прогнозирования и, следовательно, должна быть включена в средства мониторинга и бизнес-аналитики, например, в BI-платформы. Отметим, что в существующих рейтингах BI-платформ, таких как квадрант Гартнера, в 2023 г. появился критерий оценки «Интеграция с data science».

За последние годы появилось много новых методов прогнозирования временных рядов, например, STL [13], BSTS [14], Prophet [15]; широкое применения нашли методы пространства состояний, байесовской статистики и др. Возможно, их использование позволит повысить качество прогнозирования. Однако следует помнить, что нет «серебряной пули» – не метод, а качество исходных данных может обеспечить успех в прогнозировании. Заметим, что эта задача очень трудоемка (авторы статьи еще раз убедились в этом, формируя анализируемые временные ряды) и не может быть решена без разработки специальных средств парсинга данных.

Можно предположить, что использование «мягких вычислений» и нейронных сетей наряду с традиционными методами прогнозирования позволит получить более обоснованные результаты.

**Список источников**

1. Глазьев С.Ю. Теория долгосрочного технико-экономического развития. М.: ВлаДар, 1993. EDN:YSXIUV
2. Нильсен Э. Практический анализ временных рядов. Прогнозирование со статистикой и машинное обучение. СПб.: Диалектика, 2021. 544 с.
3. Хайндман Р., Атанасопулос Дж. Прогнозирование: принципы и практика. Пер. с англ. М.: ДМК Пресс, 2023. 458 с.
4. Исаев С.В., Кононов Д.Д. Исследование динамики и классификация атак на веб-сервисы корпоративной сети // Сибирский аэрокосмический журнал. 2022. Т. 23. № 4. С. 593–601. DOI:10.31772/2712-8970-2022-23-4-593-601. EDN:RUSJWB
5. Zuzčák M., Bujok P. Using honeynet data and a time series to predict the number of cyber attacks // Computer Science and Information Systems. 2021. Vol. 18. Iss. 4. PP. 1197–1217. DOI:10.2298/CSIS200715040Z
6. Ларионов К.О. Прогнозирование статистических данных атак на прикладное программное обеспечение // Проблемы современной науки и образования. 2021. № 6(163). С. 57–63. DOI:10.24411/2304-2338-2021-10606. EDN:PGVALC
7. Hobijn B., Franses P.H., Ooms M. Generalization of the KPSS-test for stationarity // Statistica Neerlandica. 2004. Vol. 58. Iss. 4. PP. 482–502. DOI:10.1111/j.1467-9574.2004.00272.x
8. Phillips P.C.B., Perron P. Testing for a Unit Root in Time Series Regression // Biometrika. 1988. Vol. 75. Iss. 2. PP. 335–346. DOI:10.1093/biomet/75.2.335. EDN:ILNEET
9. Hersbach H. Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems // Weather and Forecast. 2000. Vol. 15. Iss. 5. PP. 559–570. DOI:10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2
10. Dawid A.P., Sebastiani P. Coherent Dispersion Criteria for Optimal Experimental Design // Annals of Statistics. 1999. Vol. 27. Iss. 1. PP. 65–81.
11. Bickel P.J., Doksum K.A. An Analysis of Transformations // Journal of the American Statistical Association. 1981. Vol. 76. Iss. 374. PP. 296–311. DOI:10.2307/2287831
12. Hyndman R.J., Koehler A.B., Snyder R.D., Grose S. A state space framework for automatic forecasting using exponential smoothing methods // International Journal Forecasting. 2002. Vol. 18. Iss. 3. PP. 439–454.
13. Cleveland R.B., Cleveland W.S., McRae J.E., Terpenning I.J. STL: A Seasonal-Trend Decomposition Procedure Based on Loess // Journal of Official Statistics. 1990. Vol. 6. Iss. 1. PP. 3–33.
14. Scott S., Varian H.R. Predicting the Present with Bayesian Structural Time Series // SSRN Electronic Journal. 2014. Vol. 5. Iss. 1/2. PP. 4–23. DOI:10.1504/IJMMNO.2014.059942
15. Мастицкий С.Э. Анализ временных рядов с помощью R. 2020. URL: https://ranalytics.github.io/tsa-with-r (дата обращения 19.12.2024)

**References**

1. Glazyev S.Yu. *Theory of Long-Term Technical and Economic Development*. Moscow: VlaDar Publ.; 1993. (in Russ.) EDN:YSXIUV
2. Nielsen E. *Practical Time Series Analysis. Forecasting with Statistics and Machine Learning*. St. Petersburg: Dialektika Publ.; 2021. 544 p. (in Russ.)
3. Hyndman R.J., Athanasopoulos G. *Forecasting: principles and practice.* OTexts; 2017. 292 p.
4. Isaev S.V., Kononov D.D. A Study of Dynamics and Classification of Attacks on Corporate Network Web Services. *The Siberian Aerospace Journal*. 2022;23(4):593–601. (in Russ.) DOI:10.31772/2712-8970-2022-23-4-593-601. EDN:RUSJWB
5. Zuzčák M., Bujok P. Using honeynet data and a time series to predict the number of cyber attacks. *Computer Science and Information Systems*. 2021;18(4):1197–1217. DOI:10.2298/CSIS200715040Z

6. Larionov K.O. Forecasting Attack Statistics on Applied Software. *Problemy sovremennoi nauki i obrazovaniia*. 2021;6(163):57–63. (in Russ.) DOI:10.24411/2304-2338-2021-10606. EDN:PGVALC
7. Hobijn B., Franses P.H., Ooms M. Generalization of the KPSS-test for stationarity. *Statistica Neerlandica*. 2004;58(4): 482–502. DOI:10.1111/j.1467-9574.2004.00272.x
8. Phillips P.C.B., Perron P. Testing for a Unit Root in Time Series Regression. *Biometrika*. 1988;75(2):335–346. DOI:10.1093/biomet/75.2.335. EDN:ILNEET
9. Hersbach H. Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. *Weather and Forecast*. 2000;15(5):559–570. DOI:10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2
10. Dawid A.P., Sebastiani P. Coherent Dispersion Criteria for Optimal Experimental Design. *Annals of Statistics*. 1999; 27(1):65–81.
11. Bickel P.J., Doksum K.A. An Analysis of Transformations. *Journal of the American Statistical Association*. 1981;76(374): 296–311. DOI:10.2307/2287831
12. Hyndman R.J., Koehler A.B., Snyder R.D., Grose S. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal Forecasting*. 2002;18(3):439–454.
13. Cleveland R.B., Cleveland W.S., McRae J.E., Terpenning I.J. STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics*. 1990;6(1):3–33.
14. Scott S., Varian H.R. Predicting the Present with Bayesian Structural Time Series. *SSRN Electronic Journal*. 2014;5(1/2):4–23. DOI:10.1504/IJMMNO.2014.059942
15. Mastitsky S.E. *Time series analysis using R*. 2020. URL: https://ranalytics.github.io/tsa-with-r [Accessed 19.12.2024]

Статья поступила в редакцию 20.12.2024; одобрена после рецензирования 27.01.2025; принята к публикации 12.02.2025.
The article was submitted 20.12.2024 approved after reviewing 27.01.2025; accepted for publication 12.02.2025.

## Информация об авторах:

**НАУМОВ Владимир Николаевич** — доктор военных наук, профессор, заведующий кафедрой бизнес-информатики Северо-Западного института управления – филиала РАНХиГС
https://orcid.org/0000-0002-0385-3530

**БУЙНЕВИЧ Михаил Викторович** — доктор технических наук, профессор, профессор кафедры прикладной математики и безопасности информационных технологий Санкт-Петербургского университета ГПС МЧС России
https://orcid.org/0000-0001-8146-0022

**СИНЕЩУК Максим Юрьевич** — заместитель начальника центра информационных и коммуникационных технологий Санкт-Петербургского университета ГПС МЧС России
https://orcid.org/0009-0005-8108-3198

**ТУКМАЧЕВА Марина Алексеевна** — адъюнкт факультета подготовки кадров высшей квалификации Санкт-Петербургского университета ГПС МЧС России
https://orcid.org/0009-0004-2496-7117

Буйневич М.В. является членом редакционного совета журнала «Труды учебных заведений связи» с 2016 г., но не имеет никакого отношения к решению опубликовать эту статью. Статья прошла принятую в журнале процедуру рецензирования. Об иных конфликтах интересов авторы не заявляли.

Buinevich M.V. has been a member of the journal "Proceedings of Telecommunication Universities" Editorial Council since 2016, but has nothing to do with the decision to publish this article. The article has passed the review procedure accepted in the journal. The authors have not declared any other conflicts of interest.