

## Управление трафиком телекоммуникационных систем с использованием прогностических методов искусственного интеллекта

*Д.В. Кутузов, А.В. Осовский, Д.В. Старов, К.В. Перова*

*Астраханский государственный технический университет*

**Аннотация:** В данной статье мы рассмотрели и проанализировали различные модели прогнозирования временных рядов, используя данные, собранные с мобильных устройств IoT. Основное внимание уделено моделям, описывающим поведение трафика в телекоммуникационных системах. Рассмотрены методы прогнозирования, такие, как экспоненциальное сглаживание, линейная регрессия, авторегрессионное интегрированное скользящее среднее (ARIMA), а также N-BEATS, который использует полносвязные слои нейронной сети для прогнозирования одномерных временных рядов. В статье кратко описаны особенности каждой модели, рассмотрен процесс их обучения, и проведен сравнительный анализ качества обучения. На основе анализа данных отмечено, что для протокола UDP модель ARIMA обладает лучшим качеством обучения, для протокола TCP - линейная регрессия, а для протокола HTTPS - ARIMA.

**Ключевые слова:** телекоммуникационные системы, анализ трафика, модели прогнозирования, QoS, искусственный интеллект, линейная регрессия, ARIMA, Theta, N-BEATS.

Неуклонный рост трафика в телекоммуникационных сетях подтверждает, что общество движется в сторону мира, где данные играют ключевую роль. Эта тенденция становится естественным стимулом для развития сетей связи. Растущий рыночный спрос генерирует разнообразные технические требования, которые операторы, поставщики технологий и услуг должны постоянно удовлетворять. Постепенное отключение устаревших сетей, таких как 2G и 3G, приведет к переходу на новые технологии радиодоступа, особенно на растущих рынках, что приведет к увеличению потребления трафика данных. Индустрия мобильной связи вынуждена пересматривать традиционные стратегии в пользу новых подходов, таких, как совместное использование диапазонов спектра, сотрудничество между операторами, сети малых ячеек. Прогнозируется, что трафик, потребляемый по всему миру, будет обслуживаться не только сетями мобильной связи и 5G, но и усовершенствованными Wi-Fi-соединениями, особенно в помещениях.

---

Большая часть потребления сетевого трафика все еще связана с внутренними использованиями и сценариями с ограниченной мобильностью, такими, как дома, офисы, торговые центры и т. д. Для удовлетворения этого растущего спроса стандарты уже поддерживают конвергенцию 5G/Wi-Fi на различных уровнях, хотя массовое внедрение технических решений и связанная с ними терминальная поддержка пока не достигли заметного уровня. С увеличением числа операторов мобильной связи и увеличения скоростей доступа в Интернет объем данных трафика существенно возрос, так как нагрузка на сеть неуклонно растет. Увеличение объема трафика усложняет задачу его прогнозирования. Точное прогнозирование трафика в сотовых сетях на базовых станциях играет ключевую роль в обеспечении высокого качества услуг и является важным элементом оптимизации, проектирования и моделирования управления телекоммуникационными сетями. Исследование прогнозирования трафика в 5G и долгосрочное развитие 4G (LTE) сейчас имеют стратегическое значение для улучшения качества передачи мультимедийных данных. В последние годы прогнозирование трафика в сотовых сетях стало осуществляться с использованием глубокого обучения и статистического обучения. Адаптация технологий искусственного интеллекта для анализа и прогнозирования трафика в сетях 5G на основе сетевых данных представляет собой перспективное направление решения описанных выше задач. Эти технологии не только снизят необходимость ручного вмешательства в управление сетевым трафиком, но и обеспечат повышенную производительность, надежность и адаптивность системы, извлекая новую информацию из сетей и предсказывая условия сетевого трафика и поведение пользователей. Это, в свою очередь, позволит принимать более обоснованные решения с минимальным вмешательством человека.

---

Анализ и прогнозирование трафика представляют собой неотъемлемую часть управления современными телекоммуникационными системами и сетями. Эти прогнозы позволяют своевременно выделять и резервировать необходимые ресурсы оборудования для обеспечения достойного качества предоставляемых услуг. Два основных направления исследований посвящены проблемам точного прогнозирования трафика в проводных широкополосных сетях и сотовых сетях.

В исследованиях по прогнозированию трафика сотовых сетей применяются различные методы, включая метод экспоненциального сглаживания [1], линейную регрессию [2] и метод машин опорных векторов [3]. В работе [4] обсуждают способы повышения производительности системы QoS с использованием моделей ARIMA и NBEATS для периодических прогнозов трафика мобильной широкополосной связи.

Производительность различных моделей оценивается с использованием таких показателей, как средняя абсолютная процентная ошибка (MAPE), среднеквадратическая ошибка (RMSE) и средняя абсолютная ошибка (MAE). Результаты исследований подтверждают эффективность комбинации моделей NBEATS и ARIMA для прогнозирования трафика IP-сетей и мобильной связи.

Исследование включает в себя данные, собранные с мобильных устройств, установленных на транспортных средствах. Общий объем датасета составляет 44844 строк с 40 категорий в каждой строке. Среди этих категорий следует выделить следующие ключевые параметры:

- Index – уникальный номер транзакции;
  - Timestamp – временная метка начала транзакции (дата и время);
  - TransactionTime – продолжительность транзакции (в секундах);
  - Protocol – используемый протокол (UDP, TCP или HTTPS).
-

Два наиболее распространенных метода прогнозирования временных рядов - экспоненциальное сглаживание и модели ARIMA (модель Бокса-Дженкинса) [5] - представляют собой мощные инструменты в анализе временных данных. Модели экспоненциального сглаживания основаны на описании тренда и сезонности данных, в то время как модели ARIMA ориентированы на выявление автокорреляций во временных рядах. Структура модели ARIMA включает авторегрессионную компоненту (AR-модель), и компоненту со скользящим средним (MA-модель). ARIMA-модель  $(p, q, d)$  имеет три параметра:

$p$  - порядок авторегрессии, определяющий, насколько будущий элемент ряда зависит от предыдущих  $p$  значений;

$q$  - порядок скользящего среднего, устанавливающий связь между ошибками модели и предыдущими  $q$  значениями;

$d$  - порядок интегрирования, отражающий степень дифференцирования ряда.

Многие модели являются частными случаями [5] модели ARIMA:

- белый шум - ARIMA  $(0, 0, 0)$  без константы;
- случайное блуждание - ARIMA  $(0, 1, 0)$  без константы;
- случайное блуждание с дрейфом - ARIMA  $(0, 1, 0)$  с константой;
- авторегрессия - ARIMA  $(p, 0, 0)$ ;
- скользящее среднее - ARIMA  $(0, 0, q)$ .

Для проведения анализа данных и прогноза длительности транзакций, информация была структурирована в соответствии с протоколами: TCP, UDP, HTTPS. В данном исследовании использовались различные модели, включая наивную сезонную, экспоненциальное сглаживание, линейную регрессию, ARIMA, Theta и NBEATS. Обзор большинства этих моделей можно найти, например, в [6, с. 148 – 155]. Для реализации анализа и прогноза была выбрана библиотека Darts, специализированная для работы с

---

временными рядами в языке программирования Python. Эта библиотека поддерживает разнообразные подходы к прогнозированию временных рядов, включая как классические статистические модели, такие, как ARIMA и экспоненциальное сглаживание, так и методы, основанные на машинном и глубоком обучении. Кроме того, Darts предоставляет функции анализа статистических свойств временных рядов и оценки точности моделей прогнозирования.

Наивная сезонная модель, по своей сути, представляет собой простой метод, который повторяет несколько последних значений временного ряда. Несмотря на свою простоту, эта модель может достигать хороших результатов при удачном подборе значений.

Экспоненциальное сглаживание остается одним из наиболее распространенных подходов как для сглаживания данных, так и для прогнозирования временных рядов. Разработанный Брауном и Холтом, этот метод основан на вычислении экспоненциальных скользящих средних значений сглаживаемого ряда.

Линейная регрессия также является широко используемой моделью, предоставляющей относительно простой способ обработки данных и прогнозирования. Ее преимуществом является высокая скорость выполнения.

Theta-метод, разработанный Ассимакопулосом и Николопулосом, находит широкое применение при прогнозировании временных рядов и представляет интерес для практиков-прогнозистов. Подробности метода приведены в работе [7], а его модификация обсуждается в [8].

Метод N-BEATS [9], относящийся к методам прогнозирования с использованием нейронных сетей, представляет собой архитектуру, которая использует полносвязные слои вместо рекуррентных. Этот метод эффективен, особенно при работе с одномерными временными рядами.

---

В оценке эффективности обучения моделей использовались следующие показатели [6, 10]:

- средняя абсолютная ошибка (MAE)
- корень из средней квадратичной ошибки (RMSE)
- средняя абсолютная процентная ошибка (MAPE)
- симметричная MAPE (SMAPE).

Результаты и сравнение различных методов прогнозирования представлен ниже. Для протокола UDP сравнительные характеристики качества обучения представлены в табл. 1:

Таблица № 1

Показатели качества обучения моделей для протокола UDP

Название модели	Показатели			
	MAE	RMSE	MAPE	SMAPE
Наивная сезонная	0.8525	1.3057	54.6448	42.6230
Экспоненциальное сглаживание	0.7419	0.8947	48.3461	45.8491
Линейная регрессия	0.2580	0.5486	20.5388	13.2355
Авто ARIMA	0.2623	0.5390	19.5164	12.8306
Модель Theta	0.3015	0.5300	20.6432	15.3426
Модель N-BEATS	0.5602	0.7093	37.2326	34.0526

Для исследуемого массива данных лучшей моделью стала модель ARIMA с автоматическим подбором параметров. Результаты прогнозирования представлены на рис. 1.

Для протокола TCP сравнительные характеристики качества обучения представлены в Таблице 2.

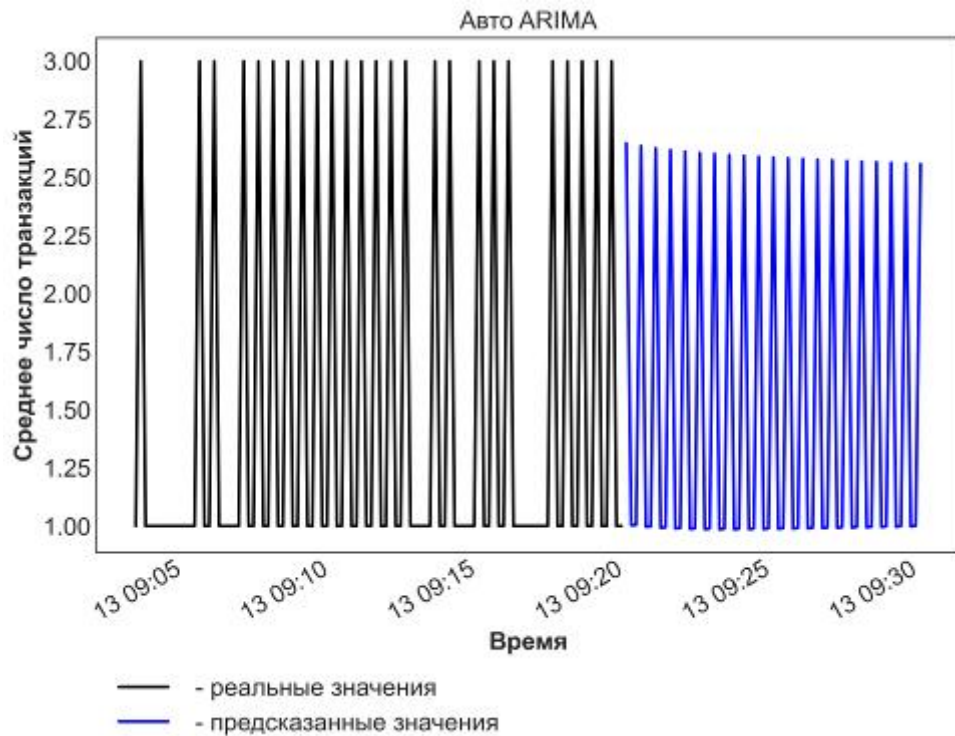


Рис. 1. – Прогнозирование UDP трафика методом ARIMA

Таблица № 2

Показатели качества обучения моделей для протокола TCP

Название модели	Показатели			
	MAE	RMSE	MAPE	SMAPE
Наивная сезонная	0.8197	1.2804	62.2951	40.9836
Экспоненциальное сглаживание	0.6289	0.8111	41.4308	40.2741
Линейная регрессия	0.3197	0.6159	25.2051	16.5640
Авто ARIMA	0.4866	0.6446	41.4338	31.1933
Модель Theta	0.3227	0.5669	23.2846	17.0089
Модель N-BEATS	0.5980	0.6970	46.6779	38.3411

Для исследуемого массива данных лучшей моделью стала линейная регрессия (рис. 2.).

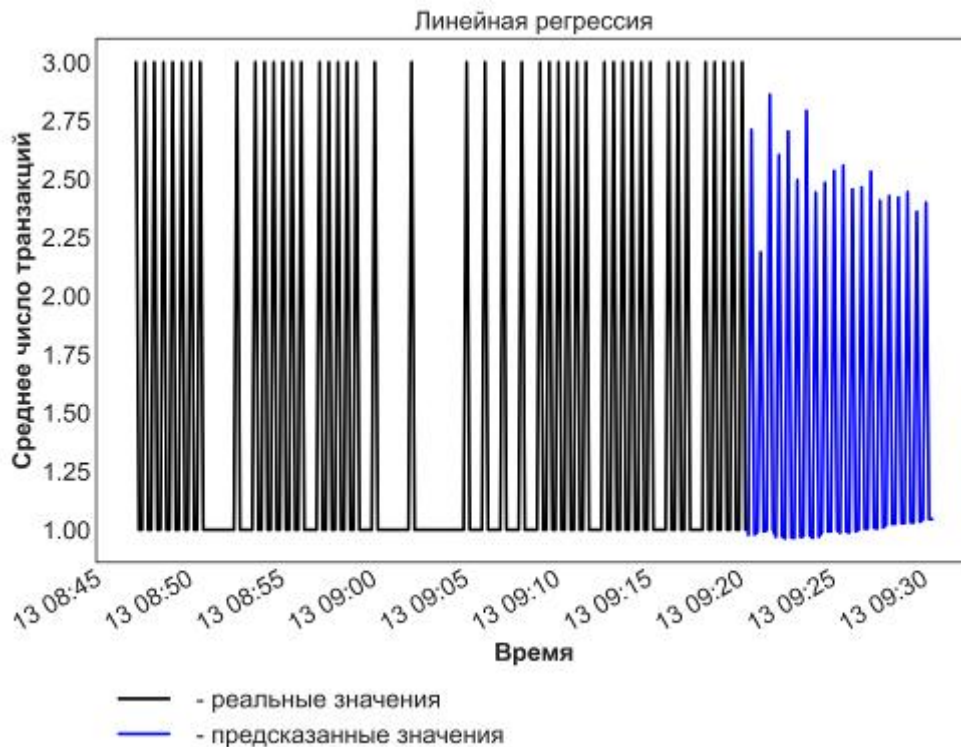


Рис. 2. – Прогнозирование UDP трафика методом ARIMA

Для протокола HTTPS сравнительные характеристики качества обучения представлены в Таблице 3:

Таблица № 3

Показатели качества обучения моделей для протокола HTTPS

Название модели	Показатели			
	MAE	RMSE	MAPE	SMAPE
Наивная сезонная	1.0820	1.4710	75.4098	54.0984
Экспоненциальное сглаживание	0.8082	0.9351	51.4153	48.3830
Линейная регрессия	0.2504	0.3788	20.4217	19.0152
Авто ARIMA	0.2375	0.4052	12.5299	11.2325
Модель Theta	0.3032	0.4311	17.9779	16.4977
Модель N-BEATS	0.5978	0.8421	31.0202	33.9271



Для исследуемого массива данных лучшей моделью стала модель ARIMA с автоматическим подбором параметров. Результаты прогнозирования представлены на рис. 3.

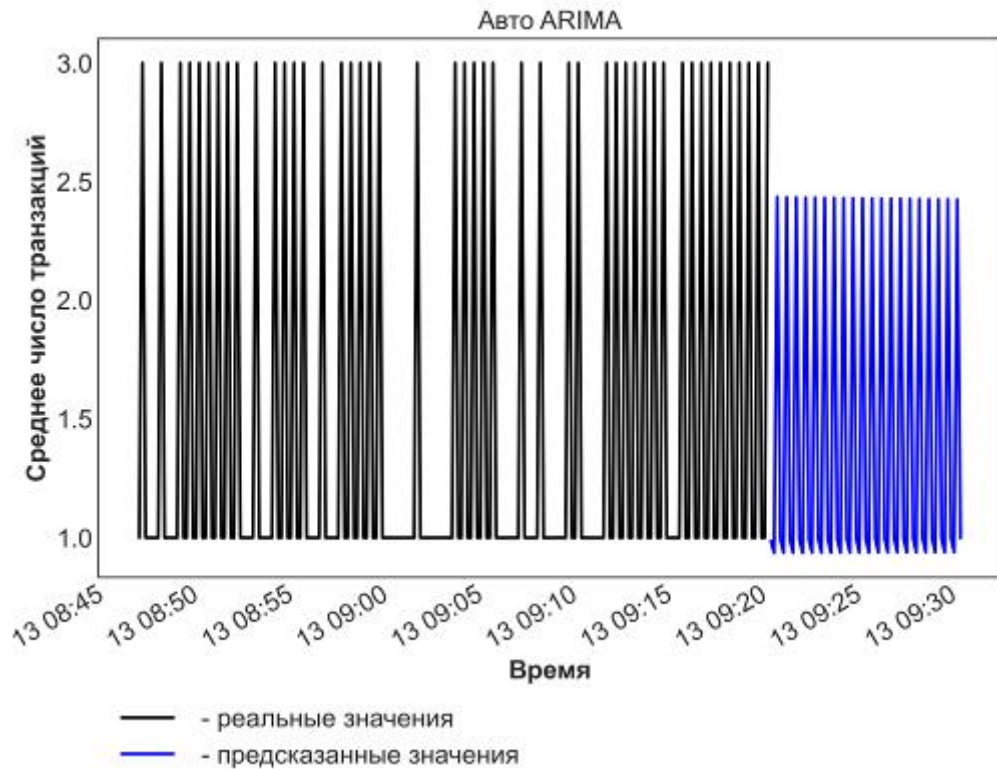


Рис. 2. – Прогнозирование HTTPS трафика методом ARIMA

### Заключение

Прогнозирование параметров трафика в современных телекоммуникационных системах играет ключевую роль в их будущем развитии и обеспечении высокого качества услуг для абонентов. Эти прогнозы не только определяют выделение ресурсов, таких, как проводные и радиоканалы, полосы частот, но и влияют на автоматическую реконфигурацию оборудования в случае сбоев или внедрения новой аппаратуры связи. В результате сравнительного анализа различных моделей обучения делаются следующие выводы: для протокола UDP наилучшей моделью для исследуемого временного ряда оказалась модель ARIMA, для

протокола TCP - линейная регрессия, для протокола HTTPS - ARIMA. Качество обучения моделей и прогнозирования зависит от характеристик исходного временного ряда, а также от качества и полноты данных, собранных для анализа. Заметим, что в ряде случаев относительно простые методы, такие, как линейная регрессия, демонстрируют приемлемое качество краткосрочного прогноза, при этом обладая значительно более высокой скоростью работы по сравнению с более сложными и точными моделями.

*Исследование выполнено при поддержке гранта РФФИ № 23-21-00196,  
<https://rscf.ru/project/23-21-00196/>*

### **Литература (References)**

1. Tran Q. T., Hao L., Trinh Q. K. J. Inf. Commun. Technol. 2010. №1. pp. 1-18.
2. Sapankevych N., Sankar R. IEEE Computational Intelligence Magazine. 2009. №2. pp. 24-38.
3. Wang J., Tang J., Xu Z. Proceedings of the IEEE INFOCOM 2017. Atlanta, GA, USA: IEEE Conference on Computer Communications, 2017.
4. Deng L., Ruan K., Chen X., Huang X., Zhu Y., Yu W. 2022 IEEE 4th International Conference on Power, Intelligent Computing and Systems (ICPICS). Shenyang, China: IEEE, 2022. pp. 336-341.
5. Hyndman R. J., Athanasopoulos G. Forecasting: Principles and Practice. Australia: Otexts. URL: [otexts.com/fpp3](http://otexts.com/fpp3)
6. Advances in Machine Learning/Deep Learning-based Technologies: Selected Papers in Honour of Professor Nikolaos G. Bourbakis. Tsihrintzis G. A., Virvou M., Jain L. C., Cham: Springer International Publishing, 2022. 231 p.
7. Assimakopoulos V. International Journal of Forecasting . 2000. №16(4). pp. 521-530.



8. Hyndman R. J., Billah B. International Journal of Forecasting. 2003. №2. pp. 287–290.

9. Oreshkin B.N., Carпов D., Chapados N., Bengio Y. N-BEATS: International Conference on Learning Representations. URL: [openreview.net/forum?id=r1ecqn4YwB](https://openreview.net/forum?id=r1ecqn4YwB)

10. Nielsen A. Practical time series analysis: prediction with statistics and machine learning, Sebastopol, CA: O'Reilly Media, Inc, 2019. 480 p.

**Дата поступления: 29.11.2023**

**Дата публикации: 3.01.2024**