

СПОСОБ ПРОГНОЗИРОВАНИЯ ДИСПЕРСИИ ЗАДЕРЖКИ ПРИ ПЕРЕДАЧЕ ГОЛОСОВЫХ ДАННЫХ

Рассмотрены существующие способы прогнозирования трафика в IP сетях. Предложен способ прогнозирования дисперсии задержки при передаче голосовых данных на основе сингулярно-спектрального анализа временного ряда. Определены критерии оценки качества способа прогнозирования. Выполнено сравнение предложенного способа прогнозирования с прогнозом на основе модели ARIMA для стационарной и мобильной сети.

Existing methods of traffic forecasting in IP networks were reviewed. The method of jitter forecasting based on singular spectrum analysis of time series is proposed. The criteria for accuracy evaluation are determined. The forecasting accuracy of the proposed method was compared with the forecasting based on ARIMA model for fixed and mobile networks.

1. Введение

Передача голосовых данных в IP сети требует соблюдения определенных ограничений, связанных с задержкой передачи и процента потери пакетов. Желательно, чтобы задержка передачи в одну сторону не превышала 150 мс, и количество теряемых пакетов не превышало 4 % [1]. В большинстве случаев эти параметры находятся в обратной зависимости, то есть чем меньше задержка пакетов может быть обеспечена, тем больше процент отбрасываемых (теряемых на конечном узле) пакетов. Обратное утверждение также справедливо. Связано это с наличием на приемном узле джиттер-буфера, который компенсирует флуктуацию задержки прибывающих пакетов.

Чем больше флуктуация задержки, тем больше размер джиттер-буфера требуется для компенсации, иначе часть пакетов будет отброшена, если пакеты придут позже времени воспроизведения. При максимальном размере джиттер-буфера появляется возможность свести количество отбрасываемых пакетов к минимуму, но при этом увеличивается время задержки. При минимальном размере джиттер-буфера время задержки уменьшается, но при этом увеличивается количество отбрасываемых пакетов.

Размер джиттер-буфера определяется значением дисперсии времени задержки прибывающих пакетов. Для сетевого трафика характерно изменение этого параметра во времени [2]. Следовательно, размер джиттер-буфера должен меняться во времени по алгоритму, учитывающему текущее состояние сети.

Существуют различные способы адаптивного регулирования джиттер-буфера [3-8]. В работе [2] предложено использовать способ на основе прогнозирования дисперсии задержки. Он позволяет поддерживать размер джиттер-буфера на минимальном уровне для компенсации флуктуации времени задержки прибывающих пакетов, обеспечивая допустимое количество их потерь.

Проблеме прогнозирования сетевого трафика посвящен ряд работ [9,10]. В них даны общие рекомендации по выбору и использованию моделей прогнозирования сетевого трафика. Недостатком этих моделей является то, что они не учитывают особенностей прогнозирования его временных параметров, следовательно, не обеспечивают требуемую точность прогноза.

Таким образом, целью настоящей работы является повышение точности прогноза временных параметров сетевого трафика за счет разработки специального способа прогнозирования.

2. Способ прогнозирования временных параметров трафика

Для построения качественного прогноза необходима разработка адекватных моделей, учитывающих особенности прогнозирования временных параметров трафика. В работе [2] показано, что временной ряд задержек передачи пакетов является процессом с медленно убывающей зависимостью (МУЗ), то есть не является стационарным. Это свойство ряда характерно для большей части случаев, наблюдаемых в IP сетях при передаче голосовых данных. Поэтому предлагается использовать мо-

дель на основе сингулярно-спектрального анализа (ССА), которая позволяет строить прогноз для нестационарных временных рядов [11].

Способ прогнозирования на основе сингулярно-спектрального анализа включает следующие действия.

1. В исходном временном ряду выделяется N последних значений:

$$\{x_i\}_{i=1}^N \quad (1)$$

2. Выбирается некоторое значение M (длина окна), такое что $M < N$, и строится траекторная матрица X по формуле 2.

$$X = (x_{ij})_{i,j=1}^{k,M} = \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_M \\ x_2 & x_3 & x_4 & \cdots & x_{M+1} \\ x_3 & x_4 & x_5 & \cdots & x_{M+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_k & x_{k+1} & x_{k+2} & \cdots & x_N \end{pmatrix} \quad (2)$$

3. Выполняется сингулярное разложение траекторной матрицы исходного ряда. Для этого определяется корреляционная матрица R по формуле 3.

$$R = \frac{1}{k} \sum_{i=1}^k \frac{1}{s_i s_j} (x_{i+l-1} - \bar{x}_i)(x_{j+l-1} - \bar{x}_j) \quad (3)$$

где:

\bar{x}_i – среднее арифметическое значение по столбцам траекторной матрицы, которое определяется по формуле 4.

s_i – стандартное отклонение, которое определяется по формуле 5.

$$\bar{x}_i = \frac{1}{k} \sum_{i=1}^k x_{i+j-1} \quad (4)$$

$$s_j = \sqrt{\frac{1}{k} \sum_{i=1}^k (x_{i+j-1} - \bar{x}_j)^2} \quad (5)$$

4. Определяются собственные числа и собственные векторы, то есть выполняется разложение корреляционной матрицы R по формуле 6.

$$R = P \Lambda P^T \quad (6)$$

где:

Λ – диагональная матрица собственных чисел

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_M \end{pmatrix}$$

• P – ортогональная матрица собственных векторов матрицы R ,

$$P = (p_1, p_2, \dots, p_M) = \begin{pmatrix} p_{11} & p_{21} & \cdots & p_{M1} \\ p_{12} & p_{22} & \cdots & p_{M2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1M} & p_{2M} & \cdots & p_{MM} \end{pmatrix}$$

5. Определяются главные компоненты Y по формуле 7 и выполняется отбор значимого количества r .

$$Y = \sum_{i=1}^k \frac{1}{s_i} (x_{i+l-1} - \bar{x}_i) P \quad (7)$$

6. Выполняется восстановление многомерного ряда \tilde{X}^* по формуле 8.

$$\tilde{X}^* = \sum_{l=1}^r y_{i,r} p_{i,r}^T \quad (8)$$

7. Выполняется денормирование матрицы \tilde{X}^* путем умножения на диагональную матрицу S , состоящую из стандартных отклонений и децентрирование путем добавления к элементам матрицы соответствующих средних значений по формуле 9.

$$\tilde{X} = \tilde{X}_0^* + \sum_{l=1}^M \tilde{X}_l^* S = \sum_{l=0}^M \tilde{X}_l^* S \quad (9)$$

В результате получена исходная матрица диагональной структуры \tilde{X} в виде суммы $(M+1)$ матриц.

8. Переход к исходному ряду может быть выполнен за счет усреднения по побочным диагоналям с помощью оператора усреднения по формуле 10.

$$\tilde{x}_s = \begin{cases} \frac{1}{s} \sum_{i=1}^s \tilde{x}_{s-i+1,i} & 1 \leq s \leq M, \\ \frac{1}{M} \sum_{i=1}^M \tilde{x}_{s-i+1,i} & M \leq s \leq k, \\ \frac{1}{N-s+1} \sum_{i=1}^{N-s+1} \tilde{x}_{k-i+1,i+s-k} & 1 \leq s \leq M. \end{cases} \quad (10)$$

9. Процедура прогнозирования на L точек для исходного ряда выполняется путем установления последнего номера элемента ряда в $N+L$ для формулы 10.

Точность предложенного способа прогнозирования определяется следующими параметрами: глубина прогноза, длина окна, выборка главных компонент. Эмпирическим способом были определены наилучшие значения этих параметров для исследуемых временных рядов: глубина прогноза – 100 точек; длина окна – по-

ловина глубины прогноза; количество главных компонент – 0,1-0,15 длины окна.

3. Экспериментальная оценка качества предложенного способа прогнозирования

Для определения качества способа прогнозирования временных параметров сетевого трафика предлагается использовать следующие критерии:

- коэффициенты переоценки K^+ и недооценки K^- , которые определяются по формулам 11 и 14 соответственно;

- средняя относительная ошибка прогноза MAPE, определяется по формуле 16.

$$K^+ = \frac{M[e_t^+]}{M[x_t]} = \frac{\sum e_t^+}{\sum x_t} \quad (11)$$

где:

e_t^+ – абсолютная ошибка прогноза, определяется по формуле 12.

$$e_t^+ = \begin{cases} e_t, & \text{если } e_t \geq 0 \\ 0, & \text{если } e_t < 0 \end{cases} \quad (12)$$

где:

e_t – абсолютная ошибка прогноза, определяется по формуле 13.

$$e_t = x_t - \hat{x}_t \quad (13)$$

где:

\hat{x}_t – значение прогноза.

$$K^- = \frac{M[e_t^-]}{M[x_t]} = \frac{\sum e_t^-}{\sum x_t} \quad (14)$$

где:

e_t^- – абсолютная ошибка прогноза, определяется по формуле 15.

$$e_t^- = \begin{cases} |e_t|, & \text{если } e_t < 0 \\ 0, & \text{если } e_t \geq 0 \end{cases} \quad (15)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|e_t|}{x_t} 100 \quad (16)$$

У лучшего способа прогнозирования временных параметров сетевого трафика должны быть минимальными и близкими по значению коэффициенты переоценки и недооценки, а также минимальной относительная ошибка прогноза.

Предложенный способ прогнозирования предлагается сравнить со способом на основе модели авторегрессии скользящего среднего – ARIMA [12], который считается одним из наиболее точных для прогнозирования сетевого трафика [13].

Модель ARIMA(p, d, q) для нестационарного временного ряда X_t представлена формулой 17.

$$\Delta^d X_t = c + \sum_{i=1}^p a_i \Delta^d X_{t-i} + \sum_{j=1}^q b_j \varepsilon_{t-j} + \varepsilon_t \quad (17)$$

где:

ε_t – стационарный временной ряд;

c, a_i, b_j – параметры модели;

Δ^d – оператор разности временного ряда порядка d ;

p – порядок авторегрессии;

q – порядок скользящего среднего.

Для построения прогноза дисперсии задержки пакетов с помощью модели ARIMA(p, d, q) необходимо определить значения параметров p, d, q , глубину и горизонт прогнозирования.

В работе [14] показано, что для прогнозирования объема сетевого трафика на 20 и более шагов вперед высокую точность показывает модель ARIMA(1,1,0). С учетом особенностей прогнозирования временных параметров трафика предлагается глубину и горизонт прогноза выбрать в 100 и 20 точек соответственно.

Для оценки качества предложенного способа прогнозирования были проведены эксперименты в условиях мобильной и стационарной сетей.

В первом случае голосовой трафик проходил через IP сеть, в которой один из сегментов был представлен системой наземной мобильной связи стандарта CDMA2000 с технологией передачи данных EV-DO Rev. A. Во втором случае голосовой трафик проходил через IP сеть, которая состояла только из стационарных наземных сегментов.

На одном из промежуточных узлов была запущена утилита tcpdump, захватывающая и сохраняющая в файл IP пакеты, проходящие через интерфейсы узла. На конечных узлах с помощью стандартных средств передачи голоса в IP сети был сгенерирован трафик. Файлы с захваченными голосовыми данными переданы для анализа в приложение Wireshark [15]. В Wireshark выделен временной ряд джиттера для всех пакетов каждого потока.

На рисунке 1 сплошной ломанной показаны значения джиттера для пакетов в стационарной сети, точками – прогноз джиттера, рассчитанный с помощью предложенного способа (прогноз ССА), пунктиром – прогноз джиттера, рассчитанный с помощью модели ARIMA. Прогноз рассчитывался в итеративном режиме с усреднением со 190 до 215 пакета с анализом 100 предыдущих пакетов на горизонт в 20 пакетов.

На рисунке 2 сплошной ломанной показаны значения джиттера для пакетов в мобильной сети, точками – прогноз джиттера, рассчитанный с помощью предложенного способа (прогноз ССА), пунктиром – прогноз джиттера, рас-

считанный с помощью модели ARIMA. Прогноз рассчитывался в итеративном режиме с усреднением со 154 до 175 пакета с анализом 100 предыдущих пакетов на горизонт в 20 пакетов.

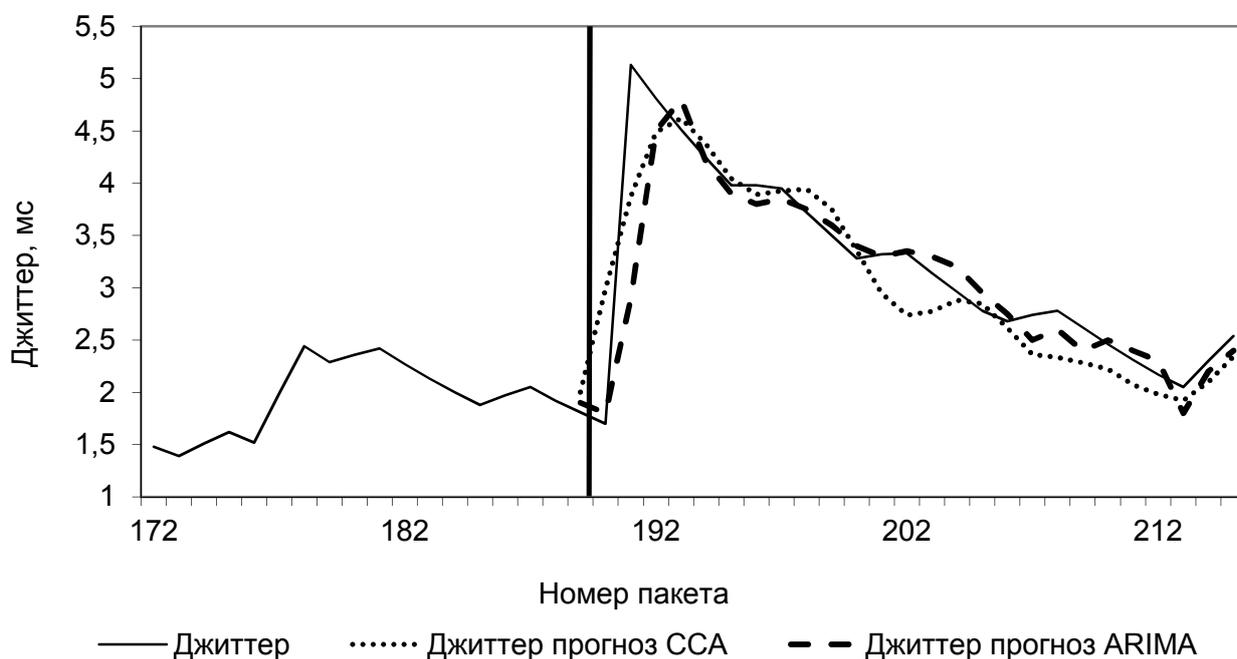


Рис. 1. Сравнение методов прогнозирования ССА и ARIMA для стационарной сети

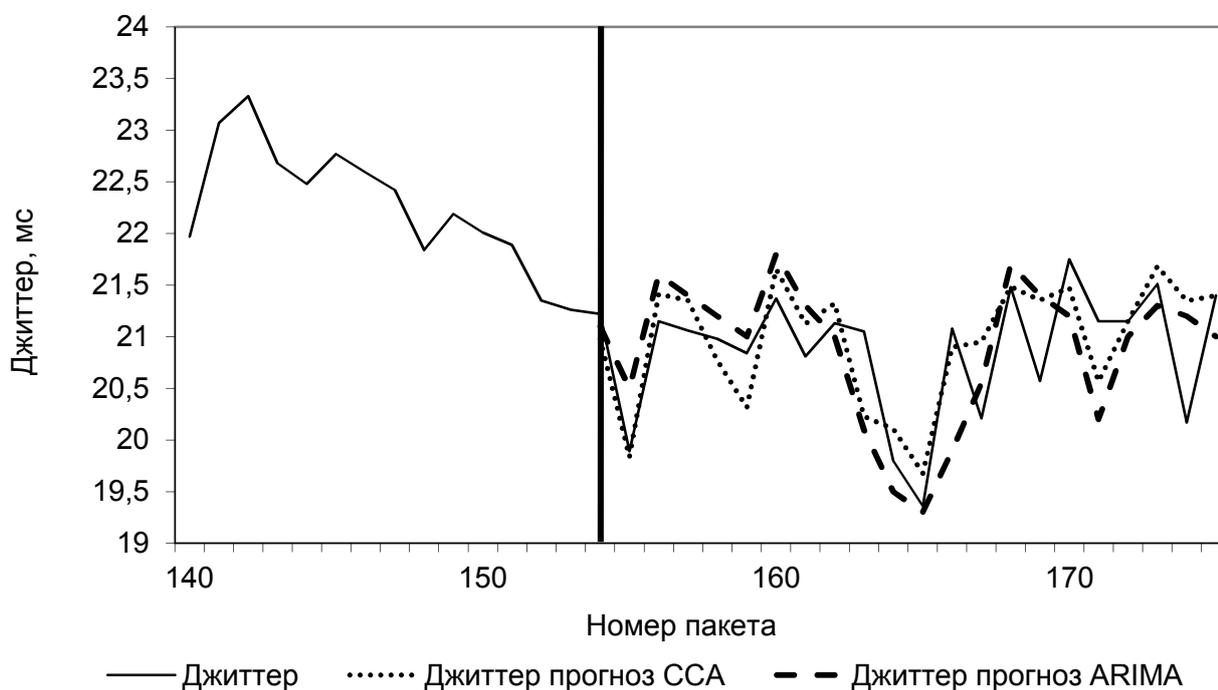


Рис. 2. Сравнение методов прогнозирования ССА и ARIMA для мобильной сети

В таблице 1 представлены коэффициенты переоценки и недооценки, а также средняя относительная ошибка для предложенного способа прогнозирования и прогноза на основе модели ARIMA для стационарной и мобильной сети.

Для стационарной сети предложенный способ прогнозирования немного уступает по точности способу на основе модели ARIMA, а в мобильной сети – превосходит его. Поэтому предложенный способ прогнозирования рекомендуется применять в мобильных сетях. В

стационарных сетях целесообразным является использование способа на основе модели АRI-МА, который требуют меньше накладных расходов на реализацию.

Таблица 1. Количественные оценки способов прогнозирования

Параметр	Стационарная сеть		Мобильная сеть	
	ARIMA	ССА	ARIMA	ССА
K^+	0,061	0,062	0,101	0,085
K^-	0,062	0,062	0,095	0,081
MAPE, %	14,3	14,5	21,8	19,5

4. Заключение

Был предложен и обоснован способ прогнозирования дисперсии задержки при передаче голосовых данных на основе сингулярно-спектрального анализа временного ряда. Эмпирически определены параметры: глубина прогноза, длина окна, выборка главных компонент, которые обеспечивают требуемую точность прогнозирования. Было выполнено сравнение точности прогнозирования предложенного спо-

соба со способом на основе модели ARIMA для стационарной и мобильной сети. По результатам сравнения даны рекомендации по применению предложенного способа прогнозирования.

В качестве дальнейших исследований необходимо разработать аналитический способ определения начальных параметров прогнозирования в зависимости от характеристик временного ряда задержек передачи пакетов.

Список литературы

1. Definition of categories of speech transmission quality, ITU-T Recommendation G.109, Jan 2007.
2. Роковой А.П. Способ адаптивного регулирования джиттер буфера в VoIP // Вісник НТУУ «КПІ». Інформатика, управління та обчислювальна техніка: збірник наукових праць - К.: "Век+", 2012.- № 55, с.199-204.
3. R. Ramjee, J. Kurose, D. Towsley, H. Schulzrinne, Adaptive playout mechanisms for packetized audio applications in wide-area networks, INFOCOM '94, Networking for Global Communications, 13th Proceedings IEEE, 1994.
4. M. Narbutt, L. Murphy, A new VoIP adaptive playout algorithm, Telecommunications Quality of Services: The Business of Success, QoS 2004, IEEE, 2004.
5. K. Fujimoto, S. Ata, M. Murata, Playout control for streaming applications by statistical delay analysis, Proceedings of IEEE International Conference on Communications (ICC), 2001.
6. C. Sreenan, J.-C. Chen, P. Agrawal, B. Narendran, Delay reduction techniques for playout buffering, IEEE Transactions on Multimedia, 2000.
7. K. Fujimoto, S. Ata, M. Murata, Adaptive playout buffer algorithm for enhancing perceived quality of streaming applications, Global Telecommunications Conference, GLOBECOM '02, IEEE, vol. 3, 2002.
8. L. Sun, E. Ifeachor, Prediction of perceived conversational speech quality and effects of playout buffer algorithms, IEEE International Conference on Communications (ICC), vol. 1, 2003.
9. Bo Zhou, Dan He, Zhili Sun, Traffic Modeling and Prediction using ARIMA/GARCH Model, Modeling and Simulation Tools for Emerging Telecommunication Networks 2006, p. 101-121.
10. Sangjoon Jung, Chonggun Kim, Younky Chung, A Prediction Method of Network Traffic Using Time Series Models, Computational Science and Its Applications - ICCSA 2006, p. 234-243.
11. Golyandina N., Nekrutkin V., and Zhigljavsky A. Analysis of Time Series Structure: SSA and Related Techniques. London: Chapman & Hall/CRC, 2001. 305 p.
12. William W.S. Wei, Time Series Analysis: Univariate and Multivariate Methods (2nd Edition), Pearson Addison Wesley, 2006.
13. B. Zhou, D. He and Z. Sun, "Traffic modeling and prediction using ARIMA/GARCH model", Modeling and Simulation Tools for Emerging Telecommunication Networks, Springer-Verlag, New York, NY, 2006.
14. G. Rutka., Network Traffic Prediction using ARIMA and Neural Networks Models// Electronics and Electrical Engineering - Kaunas: Technologija, 2008.-№ 4(84), p. 47-52.
15. <http://www.wireshark.org/>