

ПРОГНОЗИРОВАНИЕ СЕТЕВОГО ТРАФИКА ПРИ ПОМОЩИ АВТОРЕГРЕССИОННЫХ МОДЕЛЕЙ

Овчинников К.А., Бушманов В.С.

Харьковский национальный университет
радиоэлектроники, Украина.

E-mail: aymenaldulaimi@yahoo.com

Abstract

The predictability of network traffic plays a significant role in many domains such as congestion control, admission control, and network management. An accurate traffic prediction model should have the ability to capture main traffic characteristics, such as long- and short-range dependences, self-similarity etc. In this paper two models of network traffic were analyzed and it is shown that time series simple autoregressive model (AR) provides good performance on short-term predictions. It is also shown that non-stationary time series representing number of packet on router's port can be described using first-order difference ARIMA model. It was also discovered that increasing number of lag variables of AR network traffic models leads to reduction of its performance. Suggested models can be used as a part of automatic management system for prediction-based routing problem solution.

Для обеспечения надлежащего уровня надежности и качества обслуживания в современных телекоммуникационных сетях (ТКС) необходима система управления, обеспечивающая мониторинг основных параметров и реакцию на их отклонения от нормы. Процесс сбора и анализа данных мониторинга может занимать значительный промежуток времени, что приводит к снижению эффективности системы управления, вызванной запаздыванием управляющего воздействия. Для устранения указанного недостатка представляется целесообразным применение методов прогнозирования, позволяющих выявлять тенденции в изменениях параметров и осуществлять превентивное управляющее воздействие. Подобный подход аналогичен процедурам раннего обнаружения перегрузок (Random Early Detection, RED), который заключается в увеличении вероятности отбрасывания пакетов с ростом длины очереди в буфере сетевого устройства.

Данные мониторинга состояния ТКС, которые поступают в центр управления сетью, могут быть интерпретированы как временной ряд, характеризующий изменение параметров во времени. Таким образом, задача прогнозирования значений параметров состояния ТКС может быть решена стандартными средствами анализа временных рядов.

Одним из наиболее распространенных и эффективных методов анализа и прогнозирования временных рядов являются авторегрессионные модели [1,2]. Это модели, в которых значения временного ряда в некоторый момент времени линейно зависят от предыдущих значений этого же ряда. Авторегрессионная модель порядка p $AR(p)$ определяется как:

$$x(k) = c + \sum_{i=1}^p a_i \cdot x(k-i) + \xi_k, \quad (1)$$

где a_i , $i = \overline{1, p}$ – коэффициенты авторегрессии; c – некоторая константа (обычно принимается равной нулю); ξ_k – белый (обновляющий) шум; $x(k)$ – k -й элемент временного ряда.

Расширением модели (1) являются модели авторегрессии – скользящего среднего (autoregressive moving average, ARMA):

$$x(k) = c + \xi_k + \sum_{i=1}^p a_i \cdot x(k-i) + \sum_{j=1}^q b_j \cdot \xi_{k-j}, \quad (2)$$

где b_j , $j = \overline{1, q}$ – коэффициенты скользящего среднего; q – порядок модели скользящего среднего.

Ограничением для применения данных моделей является условие стационарности временного ряда, т.е. независимость его статистических характеристик от времени. Для нестационарных временных рядов используются интегрированные модели ARIMA.

$$\Delta^d x(k) = c + \xi_k + \sum_{i=1}^p a_i \Delta^d x(k-i) + \sum_{j=1}^q b_j \cdot \xi_{k-j}, \quad (3)$$

где Δ^d – разности порядка d .

Анализ интенсивности трафика на сетевых узлах (рис.1) показывает, что на длительном интервале наблюдения этот процесс может считаться стационарным. Для проверки гипотезы о стационарности были использованы тесты Лейборна-Маккейба и Дики-Фуллера с уровнем значимости 0,05 [3,4].

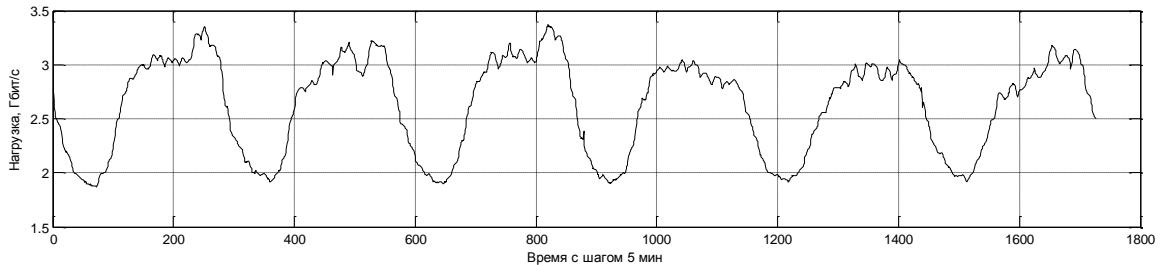


Рис. 1. Интенсивность трафика на коммутационном узле

Выбор порядка модели может быть выполнен на основании анализа вида автокорреляционной и частной автокорреляционной функции временного ряда (рис.2).

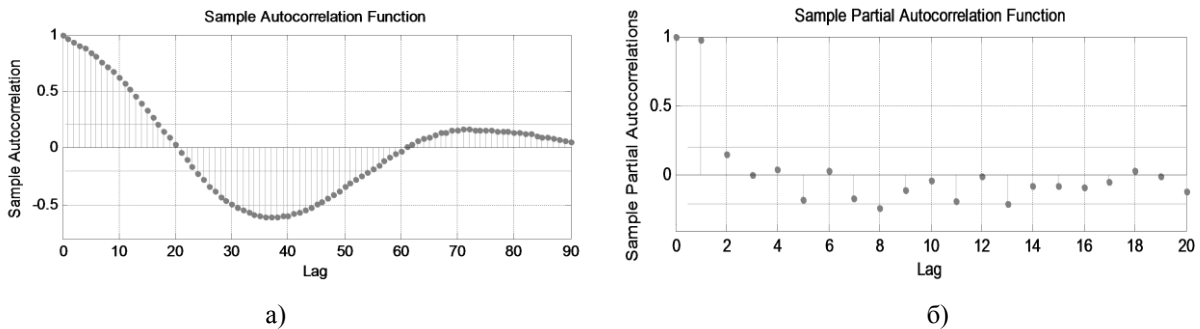


Рис. 2. Автокорреляционная а) и частная автокорреляционная б) функции временного ряда, описывающего суточные колебания интенсивности трафика

Так, согласно методике [5, 6], суточные колебания трафика могут быть описаны моделью AR(2), т.е моделью авторегрессии второго порядка. На рис.3 приведена зависимость качества прогноза от порядка авторегрессионной модели. Для оценки качества процедуры использовался критерий среднеквадратической ошибки (Mean Square Error, MSE).

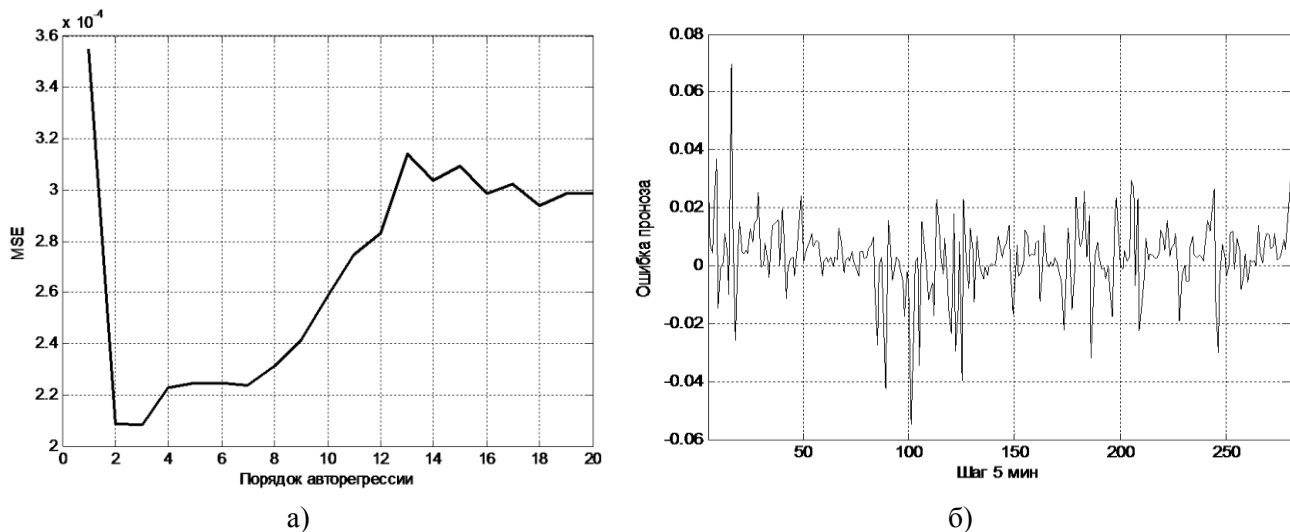


Рис. 3. Зависимость среднеквадратической ошибки прогнозирования от порядка авторегрессии p а) и значение ошибки прогноза на каждом шаге б)

Как видно из рис.3 для прогнозирования рационально использовать модель AR(2), поскольку она обеспечивает минимум среднеквадратической ошибки. При росте порядка модели авторегрессии $p > 2$ среднеквадратическая ошибка возрастает до 3×10^{-4} , что в рассматриваемом примере эквивалентно отклонению в 170 кбит/с.

При этом если рассматривать не усредненные за некоторый период времени значения интенсивности, а количество пакетов, поступающих на обслуживание за это же время, то такой временной ряд, согласно указанным критериям проверки, стационарным не является (рис.4).

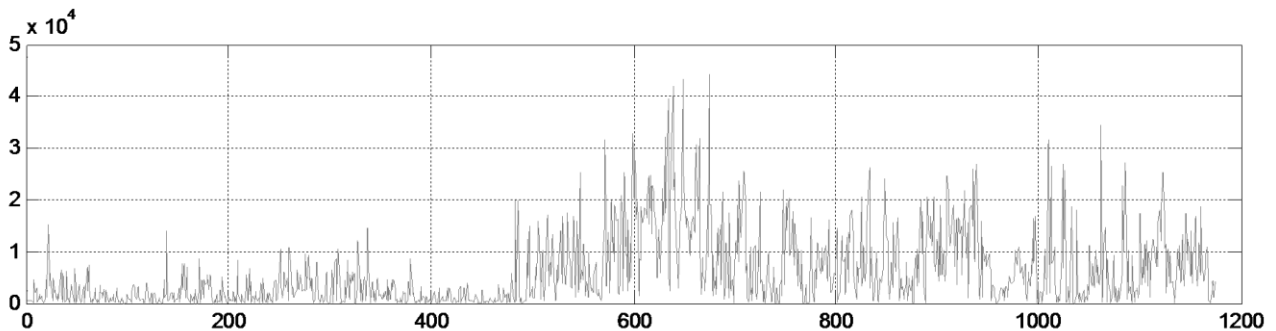


Рис. 4. Количество пакетов, поступающих на обслуживание за 5 часов с интервалом 15 секунд

Для рассмотренного случая (рис.4) стационарными с уровнем значимости 0,05 являются разности первого порядка. Анализ структуры АКФ и ЧАКФ указывает, что для описания рассматриваемого ряда достаточно модели ARIMA(2,1,0). На рис.5 представлены результаты прогнозирования количества пакетов, поступающих на вход маршрутизатора, прогноз осуществляется на один шаг.

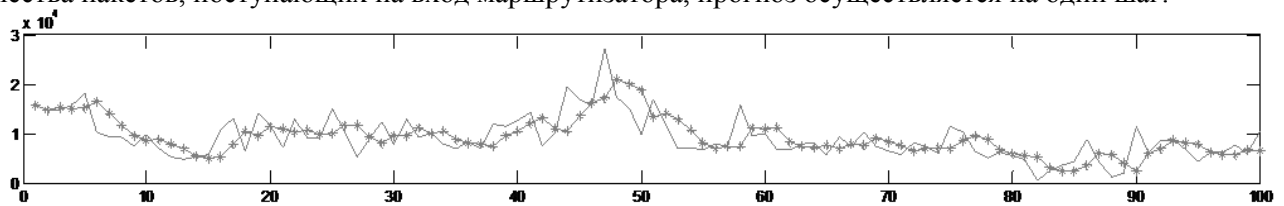


Рис. 6. Экспериментальные данные (прямая) и результаты прогнозирования (звездочка)

Таким образом, можно заключить, что задача прогнозирования интенсивности трафика может быть решена при помощи авторегрессионных моделей. При этом, если рассматриваются скоростные показатели, оптимальной производительностью обладает модель AR(2) со средней ошибкой в 170 кбит/с. Для случая, когда трафик представлен в виде совокупности пакетов, для прогнозирования необходимо применять модели ARIMA. Для рассмотренного случая модель ARIMA(2,1,0) обеспечивает среднюю ошибку на уровне 270 пакетов, при этом прогноз обеспечивает эффект сглаживания.

Литература:

1. Бокс Дж., Дженкинс Г. Анализ временных рядов, прогноз и управление: Пер. с англ. Под ред. В.Ф. Писаренко. Кн. 1. – М.: Мир, 1974. – 406 с.
2. Канторович Г. Г. Анализ временных рядов // Экономический журнал Высшей школы экономики. – 2003. – Т. 7, № 1. – С. 79-103.
3. Leybourne, S. J., McCabe B.P.M. Modified Stationarity Tests with Data-Dependent Model-Selection Rules. // Journal of Business and Economic Statistics. – 1999. – Vol. 17. – P. 264–270.
4. Dickey D.A. Fuller W.A. Distribution of the Estimators for Autoregressive Time Series with a Unit Root // Journal of the American Statistical Association. – 1979. – No 74. – p. 427 – 431.
5. Гребенников А.В., Крюков Ю.А., Чернягин Д.В. Моделирование сетевого трафика и прогнозирование с помощью модели ARIMA // Электронный журнал «Системный анализ в науке и образовании» – 2011. – №1. – Режим доступа: www.sanse.ru/download/79.
6. Агеев Д.В. Игнатенко А.А., Копылев А.Н. Методика определения параметров потоков на разных участках мультисервисной телекоммуникационной сети с учетом эффекта самоподобия [Электронный ресурс] // Проблемы телекоммуникаций. – 2011. – № 3 (5). – С. 18 – 37. – Режим доступа: http://pt.journal.kh.ua/2011/3/1/113_ageyev_method.pdf.