Autoregressive models of network traffic prediction

T. Tatarnikova¹, B. Sovetov², V. Chehanovsky³

¹Department of information technologies and security systems, Russian state hydrometeorological University, 79 Voronezhskaya ulitsa, Saint Petersburg, 195027, Russian Federation
²,³Department of information systems, Saint Petersburg Electrotechnical University "LETI", 5 ulitsa Professora Popova, 197376, Saint Petersburg, Russian Federation

tm-tatarn@yandex.ru, bysovetov@mail.ru, vvcehanovsky@mail.ru

Abstract. The relevance of the network traffic prediction is due to the requirement to store a large amount of data for a long time. In this regard, it is necessary to predict traffic volumes in order to take the necessary measures to protect and preserve data. The paper substantiates the use of autoregressive class models to construct predictive estimates. The results allowing substantiating the prospective requirements for the memory volumes of the node equipment of the infocommunication network are given.

1. Introduction

The development of infocommunication technologies every decade overcomes a new level of bandwidth limitations and leads to a number of Internet users increase.

Ease of use and accessibility of gadgets bring new media resources, as a result of which Internet traffic becomes heterogeneous. At the beginning of the XXI century, new features of network traffic, such as self-similarity and "packaging", which affect the network nodes load, were discovered.

The main property of self-similarity is memory accumulation that is a strong dependence on previous values. Self-similarity as a geometric concept emphasizes the fact that the process retains its structure over different time scales. Self-similarity as a statistical concept is characterized by the following properties: slowly decaying dispersion, long-term dependence, and the presence of a distribution with heavy "tails" of time intervals between two successive events [1].

Smoothed Internet traffic has a certain structure with a trend, which is stochastically influenced by rare bursts of packets that is referred to as packaging. Bursts of this kind affect the mathematical moments of the time sequence both on local and large time scales. Packaging can cause data loss. In this regard, it is necessary to predict the phenomenon of packaging in order to take necessary measures to protect and preserve data.

The relevance of the problem is due to the legal requirement to store a large amount of data for a long time.

The presented traffic growth forecast graphs are published on the official Cisco website with updates dated February 7, 2017. The data is related to global traffic and it is dated 2016. According to the Cisco study, mobile traffic grew 63% in 2016 compared to 2015. By the end of 2016, it had reached 7.2 EB per month, with a volume of 4.4 EB by the end of 2015. Video is expected to dominate in terms of its share in the total flow of IP traffic and overall growth in Internet traffic: 80% of all Internet traffic by 2021, while in 2016 this value was 67%. By 2021, there are expected to
be about 1.9 billion Internet video users in the world, not counting those who use exclusively mobile communications, while in 2016 there were 1.4 billion of them. By 2021, the global Internet is expected to transmit 3 trillion minutes of video per month [2]. It is found that the channel with video transmission has the highest statistical self-similarity. To this end, the forecasting demand for such an information transmission channel increases.

![Cisco forecast of traffic growth](image)

### 2. Review of existing solutions

Forecasting models are divided into two types: statistical and structural. For statistical models, the functional dependence is set analytically. These models include: regression, autoregressive and exponential smoothing. In turn, structural models are based on the dependence of structures. These include: neural network models, models based on Markov chains and decision trees. Table 1 below summarizes the advantages and disadvantages of the above mentioned methods.

| Models and methods          | Advantages                    | Disadvantages                                           |
|----------------------------|-------------------------------|---------------------------------------------------------|
| Regression                 | Modeling and design simplicity| Difficulty in finding optimal coefficients and functional dependence |
| Autoregressive             | Modeling and analysis simplicity | Impossibility to model nonlinear processes             |
| Exponential smoothing       | Modeling simplicity           | Narrow applicability of models                         |
| Neural network             | A wide variety of architectures; nonlinearity | Difficulty in choosing architectures; the size of the training sample; large time and resource-intensive training costs |
| Markov chains              | Consistent design            | Narrow applicability of models; impossibility to model long-term processes |
| Classification and regression trees | Ease of model training; scalability | Complexity of building a tree algorithm               |

In [3], statistical and structural forecasting models are analyzed, and a conclusion about the applicability of statistical models for short-term forecasting and structural models for medium-term and long-term forecasting is made. Within the considered framework of the traffic prediction problem, in order to allocate a sufficient resource for storing traffic in node equipment for a short period of time, the autoregressive model has a number of advantages over structural (trainable) models. In case of little data and no training time, this model is able to predict with less error over short time periods.
An autoregressive process is a process where the \( y \)-exponent is considered as a linear function of its previous values [4].

The AR\((p)\) model is an autoregressive model of the order \( p \), where \( p \) indicates the number of time series values included in the autoregressive equation. The AR\((p)\) model is defined as follows:

\[
y_t = c + a_1 y_{t-1} + a_2 y_{t-2} + \ldots + a_p y_{t-p} + \varepsilon_t,
\]

where \( c \) is the constant;
\( a_i \) is the \( i \)-th coefficient of a model;
\( \varepsilon_t \) is the model error.

To simplify the building of the model, a lag operator \( B^p y_t = y_{t-p} \) is introduced. Taking into account this lag operator the model (1) takes the following form:

\[
\phi_p (B) y_t = c + \varepsilon_t
\]

Let us introduce the shift operator. Then model (2) can be rewritten as:

\[
\phi_p (B) y_t = c + \varepsilon_t
\]

When building a forecast for one step ahead in AR models, the values of the latest observations are substituted into the autoregressive equation.

To build a forecast for two or more steps, the following assumption is made:

\[
y_{T+1} = 11 w_T y_T + \varepsilon_T
\]

where \( w_T \) is the forecast for one step ahead, and \( 11 w_T y_T + \varepsilon_T \) is the model (3).

The moving average model of the order \( q \) (MA\((q)\)), when building a forecast, takes into account not only the past values of the time series, but also the random error of the previous observations. This model is used to build a forecast when values of the time series have characteristic fluctuations. The MA\((q)\) model has the following form:

\[
y_t = c_1 \varepsilon_{t-1} + c_2 \varepsilon_{t-2} + \ldots + c_q \varepsilon_{t-q} + \varepsilon_t,
\]

where \( q \) is the order of the MA\((q)\) model, denoting the number of previous values of random deviations.

When substituting the lag operator \( B \) into expression (5) and deducing \( \varepsilon_t \) from parentheses, we obtain:

\[
y_t = (1 + c_1 B + c_2 B^2 + \ldots + c_q B^q) \varepsilon_t.
\]

For a more concise notation, we introduce an operator, and then the MA\((q)\) model can be defined as follows:

\[
\theta_q (B) y_t = c_1 B + c_2 B^2 + \ldots + c_q B^q \varepsilon_t.
\]

To build a forecast for several steps, a forecast for one step is calculated and it is assumed that \( y_{T+1} = \tilde{y}_{T+1} \).

The processes AR\((p)\) and MA\((q)\) are interconnected and can be described by the ARMA\((p, q)\) model:

\[
y_t = c + a_1 y_{t-1} + a_2 y_{t-2} + \ldots + a_p y_{t-p} + c_1 \varepsilon_{t-1} + c_2 \varepsilon_{t-2} + \ldots + c_q \varepsilon_{t-q} + \varepsilon_t
\]

and using the operators \( \phi_p \) and \( \theta_q \), the ARMA\((p, q)\) model can be represented as follows:

\[
\phi_p (B) y_t = \theta_q (B) \varepsilon_t + c.
\]
Obviously, this model combines the properties of the autoregressive and moving average models, since on the one hand, the predicted values of the model correspond to the values of AR\((p)\), and on the other hand, considering MA\((q)\) makes it possible to more accurately approximate the time series and cut off unnecessary autoregressive values.

An important condition for building AR\((p)\), MA \((q)\), and ARMA \((p, q)\) models is the condition of stationarity of the time series, i.e. its basic properties, such as mathematical expectation, variance, autocorrelation coefficients, remain unchanged over time.

When a non-stationary time series is observed, it is necessary either to use other models, or to bring the series to a stationary form.

To bring the series to the stationary form, the difference operator \(\Delta^d y_i = (1 - B)^d y_i\) is used, where \(d\) is the number of successive subtraction operations of the time series levels, which ensures the series stationarity.

The ARIMA \((p, d, q)\) model is an autoregressive and integrated moving average model, which is built on the basis of differences, and it has the following form:

\[
\phi_p(B)(1 - B)^d y_i = \theta_q(B)e_i + c.
\]

It can be noted that each subsequent model becomes more complicated and is a certain evolution of the previous one. ARIMA, from a modeling point of view, constitutes a broad and flexible class of stochastic models. Therefore, in this work, the ARIMA model is used to predict network traffic.

When studying time series using the ARIMA model, it is necessary to choose the parameters \(p, d, q\). The choice of these parameters determines the accuracy of the process description and, consequently, the accuracy of the prediction. To select the parameters \(p, d, q\), autocorrelation analysis is used.

The \(p\) parameter is selected from a partial autocorrelation model, which allows measuring the relationship between the current level of the time series and its previous values.

The parameter \(d\) is chosen based on the number of performed consecutive subtraction operations. As a rule, \(d\) is limited to the number \(d = 2\), since taking the second differences makes it possible to reduce any non-stationary series to a stationary form.

The parameter \(q\) is selected from the autocorrelation model, which allows measuring the dependence between the function and its shifted copy values on the time shift value.

3. Real data forecasting

To obtain medium and long-term forecasts, long records of traffic traces are required [5]. This seemingly impossible condition has been fulfilled thanks to the open publication of Internet traffic data by Japan's MAWI research group of the WIDE project. The training set includes 24-hour tracks for 09/05/2018 and 09/04/2019, 48-hour tracks for 09-11/01/2007, 72-hour tracks for 03/18-20/2008, 96-hour tracks for 03/30/2009 and 02/01-04/2009.

Fig. 2 shows examples of the number of transmitted bits for two days on 04/09/2019 and 04/10/2019 (48-hour track consists of 192 counts with 15 minutes time step). The y-axis shows the amount of the transmitted traffic, and the x-axis shows 15 minutes time steps.

The measure of self-similarity is estimated by the Hurst parameter \(H\). In this study, the estimate of \(H\) is performed by the method of the normalized range R/S:

\[
H = \frac{\log \left( \frac{R}{S} \right)}{\log \left( \frac{N}{2} \right)},
\]

where \(N\) is the length of the time series;
\(R\) is the range;
\(S\) is the standard deviation.
If $H > 0.5$, then the process is self-similar.

![Fig. 2. Changing of the transmitted bits amount](image)

Fig. 2 shows a graph of the transmitted bits amount $V$, GB.

![Fig. 3. R/S ratio](image)

Fig. 3 shows a graph of the normalized range R/S, from which the value $H = 0.6633$ is obtained.

Thus, a real sample of the main parameters of Internet traffic is obtained and statistical tests for self-similarity are carried out. The considered time series are self-similar.

The ARIMA model is applied to the obtained data of real network traffic. To determine the optimal parameters of the models, a full enumeration is carried out for all possible values. The best parameter values of the ARIMA model $(p, d, q)$ are:

- $p$, which is the order of lag, equals 3;
- $d$, which is the difference operator power, equals 1;
- $q$, which is the power of the moving average, equals 2.

The division of data into training and test data is reduced to a ratio of 3 to 1. Initial data is normalized, i.e. the entire numerical sequence is divided by the maximum values from the series. MSE (Mean Squared Error) is chosen as the main criterion for the deviation from the initial data:

$$MSE = \frac{1}{n} \sum_{j=1}^{n} (Y_j - \hat{Y}_j)^2,$$

where $Y_j$ is the original value,

$\hat{Y}_j$ is the predicted value.

Fig. 4 shows a comparison between the original traffic and traffic predicted by the ARIMA model. The MSE value of 0.004 indicates a high degree of efficiency of ARIMA model application for the prediction of the self-similar traffic.
From the prediction results presented above, it can be concluded that the models repeat the geometry with a slight deviation.

Autoregressive class models are quite suitable for integration into network devices as an analytical program working on aggregate data, which allows real-time prediction of traffic volume behavior over a short period of time.

### 4. Conclusion

The growth of the heterogeneous traffic volumes in infocommunication networks actualizes tasks of ensuring the quality of the provided communication services, which in turn requires the forecasting models usage. The results obtained in the work allow substantiating the prospective requirements for the memory volumes of the node equipment of the infocommunication network.

### References

[1] Data Growth, Business Opportunities, and the IT Imperatives, available at: https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm

[2] Electronic text data, available at: https://www.cisco.com/c/ru_ru/about/press/press-releases/2017/06-09b.html.

[3] Sovetov B. Ya., Tatarnikova T. M., Poymanova E. D. 2019 Organization of multi-level data storage. Information and Control Systems [Informatsionno-upravliaiushchie sistemy – in Russian] 268. doi:10.31799/1684-8853-2019-2-68-75

[4] Lorido-Botran Tania, Miguel-Alonso Jose, Lozano Jose A. 2014 A review of auto-scaling techniques for elastic applications in cloud environments. Journal of grid computing. 12(4) 559.

[5] Kutuzov O. I., Tatarnikova T. M. 2018 Model of a self-similar traffic generator and evaluation of buffer storage for classical and fractal queuing system. Moscow Workshop on Electronic and Networking Technologies, MWENT, 2018.