UDC 681.3.06

L.A. Tereikovska, PhD, Assis. Prof. (Kyiv National University of Construction and Architecture, Ukraine),
K.O. Radchenko, Assis. (NTUU "Igor Sikorsky KPI", Ukraine)
K.O.Kharlamov, student (Kyiv National University of Construction and Architecture, Ukraine)

The application of discrete wavelet transform for forecasting the operational lo ad of the Web server of the distance education system

Web server of the system of distance learning. Using the theory of discrete wavelet transforms, the mathematical support of such models is developed. The ways of improvement of models are indicated.

Forecasting the operational load of the Web server.

At present, the widespread introduction of distance education systems in educational institutions has taken the form of a pronounced and stable trend. In most higher education institutions distance education is conducted using such specialized platforms as: Lotus LearningSpace, Прометей, WebCT CE, WebCT Vista, Blackboard, ATutor, OLAT and Moodle. The domestic universities mainly use the Moodle platform. Regardless of the type of platform, the technical basis for its operation is a Web server that serves requests from remote browser clients. Note that the Moodle platform uses Apache as its Web server, which in the standard configuration are not able to simultaneously serve more than 30 clients. Thus, there arises the task of forecasting the load on the Web server, which can be solved by developing appropriate models [3]. In connection with the fact that the process of connecting remote clients to the distance education system has a clearly expressed random character, then, according to a number of literature sources [3-5], it is expedient to use statistical models to forecast the load.

Obviously, the type and parameters of the forecast model depend to a large extent on those parameters that actually determine the load on the Web server. Based on the typical tasks performed by the Web server, as well as the data transfer protocols used, the load can be estimated from the groups of parameters that characterize the use of:

1. Hardware resources of computer-server – RAM, hard disk, microprocessor.
2. Operating system resources - the length of the network connection queue, the number of open files, the number of running processes, etc.
3. Network resources - the volume of incoming and outgoing traffic, the size of network packets, the number of clients per unit of time, the number of "improper" network packets for each of the used network protocols, etc.

To predict the values of these parameters in time, such statistical models of data prediction as linear (1) and polynomial (2), exponential (3), moving average (4) and exponential smoothing (5) are mainly used.

$$P = a + b \times t \quad (1),$$

$$P = \sum_{i=0}^{K} \left(a_i \times t^i \right) \quad (2),$$

$$P = a \times e^{k \times t} \quad (3),$$

$$P_{t+1} = \frac{P_t + P_{t-1} + \ldots + P_{t-n+1}}{n} \quad (4),$$

$$S_{t+1} = \alpha \times P_t + (1 - \alpha) \times S_{t-1} \quad (5),$$

where P is the parameter value, t is the time, a, b are the linear regression coefficients, K is the degree of the approximating polynomial, a_i is the i-th coefficient of the approximating polynomial, k is the coefficient, P_t is the parameter value at time t, n is the size of the sliding window (defined in time readings), S_t is the smoothed value of the parameter at time t, and α is the exponential smoothing coefficient.

As an example of typical dependencies of parameters on time, the graphs shown in Fig. 1 and Fig. 2.

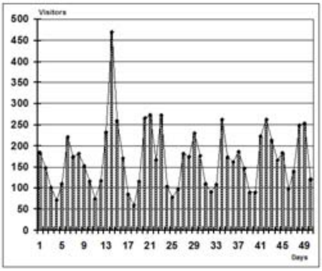

Fig. 1. Dependence of Web site attendance on time

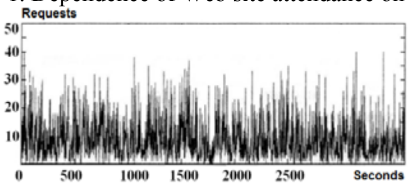

Fig. 2. Dependence of the number of requests to the Web server on time

In Fig. 1 shows the graph and the histogram of the number of visitors to a Web site of one of the domestic universities for 50 days using the registration window for 1 day. In Fig. 2 shows the graph, given in [3], of the dependence of requests to the Web site on time in seconds, when using the registration window.

One of the most modern methods of time-frequency analysis of non-stationary processes is the wavelet transform. In [3] a model is developed that allows calculating the frequency-time characteristics of Internet servers. However, this model is based on the theory of continuous wavelet transform, and therefore has significant redundancy and requires adaptation to discrete data, which corresponds to typical statistics of Web server parameters [1, 2].

A possible way to eliminate these drawbacks can be the use of a discrete wavelet transform. Thus, the main purpose of this study is to develop a wavelet model for predicting the operational parameters of Web servers of distance learning systems

In the general case, the continuous wavelet transform of a function *f(t)* with finite energy in the space $L^2(R)$ is written as follows

$$W(a,b)=|a|^{-0,5}\int_{-\infty}^{\infty} f(t)\psi^*\left(\frac{t-b}{a}\right)dt \qquad (6),$$

where W is the wavelet transform coefficient, ψ is the base wavelet (basis function), * is the complex conjugation procedure, a is the wavelet scale, b is the wavelet shift, $a,b\in R$, $a\neq 0$.

Moreover, the function ψ must satisfy the following requirements: its zero moment must be equal to 0, the energy of the function must be finite, be concentrated within a finite interval, and rapidly decrease to zero outside this interval [1, 2]. To analyze the series with a polynomial trend, the central moments of the v-th order must vanish in the basis wavelets.

A feature of the discrete wavelet transform of a continuous function is the use of discrete scale values and wavelet shift. As a rule, these quantities are given in the form of power functions of the form

$$a = a_0^{-m} \qquad (7),$$

$$b = k \times a_0^{-m} \qquad (8),$$

where m is the scale parameter, k is the shift parameter, a_0 is the initial scale, m, k, a_0 are integers, and $a_0 > 1$.

Taking (7, 8) into account, we write expression (6) in this way

$$W(m,k)=|a_0|^{0,5m}\int_{-\infty}^{\infty} f(t)\psi^*\left(a_0^m \times t-k\right)dt \qquad (9),$$

Quite often, a_0 is assumed to be equal to 2. Such a discrete wavelet transformation is called a dyadic transformation [2]. For the dyad wavelet transform, the expressions (7, 8, 9) are transformed so

$$a = 2^{-m} \qquad (10),$$

$$b = k \times 2^{-m} \qquad (11),$$

$$W(m,k)=2^{0,5m}\int_{-\infty}^{\infty} f(t)\psi^*\left(2^m \times t-k\right)dt \qquad (12),$$

From the point of view of discrete wavelet analysis, the statistics of the Web server have the following features:

- Limited by the interval $t \in [t_{min}, t_{max}]$.

- Data logging is carried out with a certain discreteness Δt.

- For a dyadic transformation, the number of points of the N series must be equal to

$$N = 2^z \quad (13),$$

where z is an integer.

Taking these features into account, expressions (9, 12) are transformed in this way

$$W(m,k) = a_0^{\,0,5m} \sum_{i=1}^{N} \left(f(t_i) \psi^*(a_0^m \times t_i - k) \right) \quad (14),$$

$$W(m,k) = 2^{0,5m} \sum_{i=1}^{N} \left(f(t_i) \psi^*(2^m \times t_i - k) \right) \quad (15),$$

where N is the number of points in the series, $f(t_i)$ is the value of the data series at time t_i (the i-th registration moment).

The inverse discrete wavelet transform, which in its essence is the restoration of a function by a set of wavelet coefficients, is defined by the expression:

$$f = \frac{\pi}{\ln a_0} \sum_{m=0}^{N-1} \sum_{k=0}^{L-1} \psi^* W_{m.k} \quad (16),$$

where L is the number of scales.

In order to verify the models (7-16) on real data, a dyad wavelet analysis of attendance statistics of the Web site was carried out, the graph of which is shown in Fig. 1. Taking into account the need to perform (13), only the first 32 points of the series were involved in the analysis. In this case, $z = 5$, and $m_{max} = 4$. The basic wavelet of Daubechis is used. Calculations were conducted using the Mathcad package. As a result of the analysis, the wavelet spectrum curves shown in Fig. 3 are obtained.

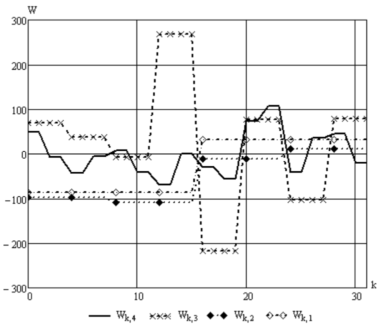

Fig. 3. Wavelet spectrum of attendance of the Web server

Analysis of the data in Fig. 3 indicates that in the process being investigated, there are 4 periodic components that are equal to 2, 4, 8 and 16 days. You can also notice that the third component occurs approximately on the 14-16th day of operation, and the rest do not have a clear localization in time. The received results testify to the prospect of calculating the frequency-time characteristics of the attendance of the Web server with the help of discrete wavelets.

An inverse discrete wavelet transform of the signal is also performed on a set of wavelet coefficients. The graphs of the reconstructed and original signal are shown in Fig. 4.

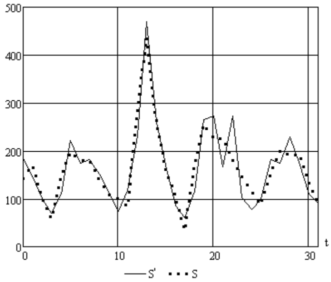

Fig. 4. Graphs of source and recovered signal

The sufficiently high similarity of the indicated graphs confirms the possibility of applying this approach and the adequacy of the developed mathematical models. However, additional studies are required to use the results obtained in the technical condition forecast model.

Conclusions.

1. Using the methods of the theory of discrete wavelet transforms, a mathematical model for calculating the frequency-time characteristics of operational parameters of Web servers of distance learning systems is developed.

2. Prospects for further research in this direction are to develop a technique for optimizing the form of the basis wavelets.

References

1. Астафьева Н.М. Вейвлет-анализ: Основы теории и примеры применения / Н.М. Астафьева // Успехи физических наук. – 1996. – т.166, № 11 – С. 1145–1170.
2. Дьяконов В.П.Вейвлеты. От теории к практиве / В.П. Дьяконов. – М. :Солон, 2006. – 440 с.
3. Менаске Д. Производительность Web-служб. Анализ, оценка и планирование / Менаске Д., Виргилио А. ; пер. с англ. – СПб.: ДиаСофтЮп, 2003. – 480 с.
4. Михальов О.І. Вейвлет-мультифакторний аналіз складних зображень / О.І. Михальов, Ю.О. Водолазький // Вісник Вінницького політехнічного інституту. – 2009. – № 2 – С. 84–87.
5. Терейковская Л. А. Применение дискретного вейвлет-преобразования для определения частотно-временных параметров эксплуатационной нагрузки Web-сервера системы дистанционного образования [Текст] / Л. А. Терейковская // Науково-технічний збірник «Управління розвитком складних систем» КНУБА. – 2011. – Вип. 5. – С. 124 – 128.