JOURNAL OF APPLIED
COMPUTER SCIENCE
Vol. 18 No. 2 (2010), pp. 91-101

Packet Switching Networks Traffic Prediction Based on Radial Basis Function Neural Network

Arkadiusz Zaleski, Tomasz Kacprzak

Institute of Electronics, Technical University of Łódź
Wólczańska 211/215, 90-924 Łódź, Poland
azaleski@eranet.pl, tomasz.kacprzak@p.lodz.pl

Abstract. *New multimedia applications require Quality of Service support, which is still not successfully implemented in current packet-switched networks implementations. This paper presents a concept of neural network predictor, suitable for prediction of short-term values of traffic volume generated by end user. The architecture is Radial Basis Function neural network, optimized with respect to a number of neurons. Testing mode of the neural network is very fast, what enables application of this tool in nodes of telecommunication network. This would help to warn a network management system on early symptoms of congestion expected in the near future and avoid the network overload.*
Keywords: *telecommunication traffic, congestion control, neural networks, time series prediction, QoS, traffic flow.*

1. Introduction

Despite the enormous development of telecommunication networks, the problem of network congestion is still a critical issue. New broadband access technologies (ADSL2+, WiFi, UMTS-HSDPA, cable TV modems) increase network traffic rapidly, what generates demand for speed and bandwidth in packet-switched networks: ATM, current and future TCP/IP Internet, frame relay, etc., [1-7]. New

Internet applications and integrated multimedia services require QoS support (service and data protection), because global models like IntServ and DiffServ are still not implemented. Currently used techniques like TCP mechanisms are not effective enough in many cases, so hardware extension is still basic solution for network overload prevention.

The congestion is defined as a state of network elements in which they are not able to meet the performance objectives for the already established connections and/or for the new connection requests, [7]. In case of networks based on TCP/IP protocols stack, the congestion control is implemented as an open-loop, dynamic window scheme. This is an example of so called reactive scheme, because most of TCP mechanisms (RTT variance estimation, exponential backoff, Karn's rule, slow-start, dynamic window sizing, fast retransmit and fast recovery) starts after the congestion has occurred and therefore its effectiveness is not satisfactory. This means that some transferred data have been already lost or their delivery is delayed, what is not acceptable for some applications. The only Explicit Congestion Notification mechanism for transport protocols would be very useful for network overload prevention, but it is not included in all TCP implementations and routers, so it has to be negotiated at connection establishment. The other solutions like Stream Control Transmission Protocol [8] or measurement based traffic shaping with ICMP protocol support have also some serious disadvantages. SCTP was developed mainly to eliminate impact of head-of-line blocking effect on transport protocol performance, because in case of TCP protocol it causes messages transmission delay, what is not acceptable e.g. for signalling. Measurement based congestion control reduces network utilization and ICMP protocol is often used for Denial of Service attacks purpose like e.g. flooding.

A better approach is to apply a control system, that should be preventive in order to react quickly before the excessive load reaches saturation level, thus resulting in avoidance as well as minimization of the spread of congestion. However, this method requires reliable short-term traffic forecasting, which could serve as a tool for sensing the early symptoms of approaching congestion state.

The idea of traffic load forecasting is based on the assumption that similar state at the near future moment have existed in the past. This results in a model of time series, where measured traffic parameter values for contiguous time intervals up to the actual time point are used to predict the value for next time interval.

In this paper we propose neural network forecasting model, suitable for prediction of short-term values of end user traffic descriptor parameters e.g. average or maximum bit rate for particular flow within assumed time interval. The architec-

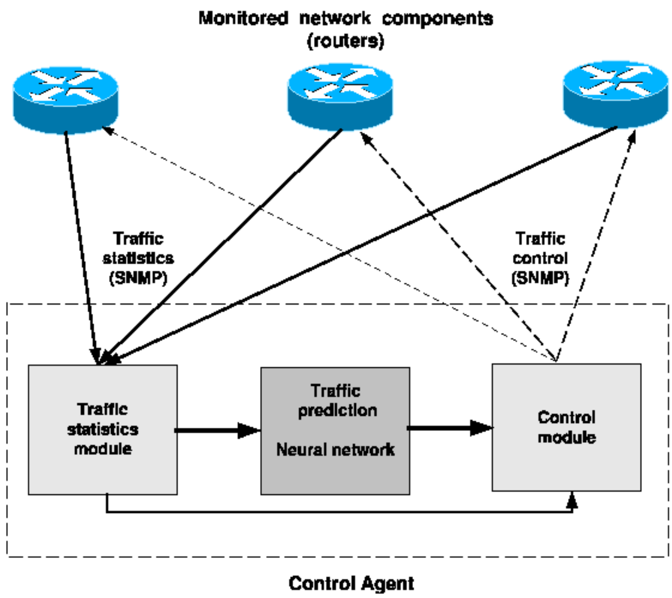

Figure 1. Intelligent agent-based system for congestion control

ture is Radial Basis Function neural network, optimized with respect to a number of neurons. Testing mode (prediction process) of the neural network is very fast, which enables application of this tool in a broadband telecommunication network.

2. Agent-based system for congestion avoidance

Despite the vast research efforts and the large number of different congestion recovery schemes proposed, there is so far no universally acceptable one. In this paper we propose an intelligent agent-based structure for traffic prediction, which could help to avoid congestion in packet-switched telecommunication networks. The agent could be implemented both in network nodes as well as a separate application server connected via SNMP with a group of served network elements (hosts, routers, switches, etc.) and its activity is limited to nodes belonging to the specified cluster. The agent structure is composed of three blocks: Traffic statistics module, Control module and Neural network traffic predictor, Fig. 1.

The module of traffic statistics monitors on-line flows of traffic passing the observation point (typically it is node) and gathers necessary information of their long-term behaviour in the past. This information is next prepared in the form of input data for neural network traffic predictor, so the quality and proper selection of statistics for particular traffic parameter have direct impact on speed of training process and prediction precision. The neural network is a time series predictor using lag points from a sequence of real-time values in the past. The output of this block is the predicted value of specified traffic parameter, expected for next time interval. Based on this value and additional information the control module can prepare a decision, which influences traffic flows in the particular router. This decision can result in, for example, changing routing procedure, controlling sliding window or retransmission time out.

3. Neural network for time series prediction

The time series is a set of ordered data points $V(t)$ spaced typically with uniform time intervals, where $t = 0, 1, 2, 3, ...$ symbolizes elapsed time. In analyzed case mentioned data points are represented by selected traffic parameter measurements values. The goal of time series analysis is to detect a nature of phenomenon, which is described by a sequence of measurement data taken from the past, to predict future values of this time series. Both tasks require identification and formal description for time series components, so that fixed pattern could be applied again for any other data within the same model.

There are a lot of time series analysis methods like spectral analysis, exponential smoothing or autoregressive integrated moving average (ARIMA), [9]. Unfortunately, mentioned techniques can't be used for packet traffic parameters prediction, due to their bursty and fast time-varying characteristics and huge amount of flows processed in core network routers. Traditional techniques provide optimal solutions only in case of stable states, because of real time calculations requirement and the prediction system should be able to process a lot of data in real time.

Lately, more and more popular are methods based on statistics and artificial intelligence like e.g. genetic algorithms or artificial neural networks, which are already used for many other applications like function approximation, modelling and pattern recognition, [10-12]. Using neural network for time series prediction purpose instead of traditional techniques, could reduce errors which descend from modelling, approximation and unpredictable changes in the monitored system. Additionally, they are able to transmit relationships discovered during training process to new cases, which were not used during the training stage (generalization). Neu-

Algorithm / Network	Elman	Perceptron	Radial (self-organizing)
Gradient descent	2091	6532	
Conjugate gradient	65	100	
Quasi-Newton BFGS	17	29	4
Levenberg – Marquardt	3	7	
RPROP	84	369	

Figure 2. Efficiency comparison for selected networks and training algorithms

ral network scheme trained by sufficiently long time series taken from the history should be a good solution for prediction of one point ahead and it was also proposed in [12], where multilayer perceptron has been applied to predict 24-hour load shape of Polish power system. The objectives of our research are to confirm if neural network approach is a good solution for packet traffic parameters prediction and to determine the optimal neural network model for this purpose.

This task requires making decision on neural network type (architecture), training algorithm and training data set. Efficiency research for different solutions has to return the optimal architecture, with number of training epochs, input data vector size and number of hidden layers neurons reduced to minimum. It is effective on number of floating-point operations (processor capacity) demand, what could be important in case of large systems handling, when traffic prediction for many users/flows has to be performed within limited time period.

For time series prediction task three types of neural network are recommended: feedforward neural networks e.g. multilayer perceptron, recurrent networks e.g. Elman network and Radial Basis Function neural network [12]. These three kinds of architectures had been extensively tested for the same training data set and their complexity had been compared as well as number of training epochs required to obtain an assumed mean square error value. The efficiency comparison result for example time series prediction task was presented in Fig. 2, where the number of training epochs required to reduce training error to the assumed value was calculated.

The Radial Basis Function (RBF) neural network of the architecture given in Fig. 3 has been chosen as the best solution from these three ones, due to good generalization and quick training process. The RBF network used has three layers.

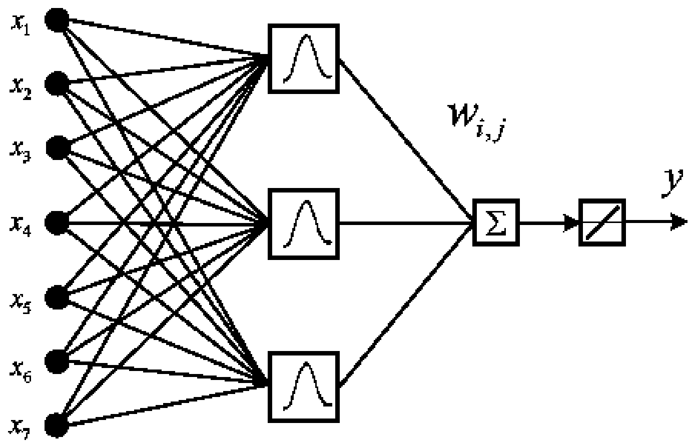

Figure 3. Architecture of Radial Basis Function neural network under test

At the input layer 7 neurons act as distributing points. At the hidden layer 3 neurons use the Gaussian-like response function, while at the output layer one neuron acts as a linear summer of hidden layer weighted outputs.

The radial transfer function is of the form:

$$\phi(\|\bar{x} - \bar{c}\|) = exp\left(- \frac{\|\bar{x} - \bar{c}\|^2}{2\alpha^2} \right) \tag{1}$$

where: $\bar{x}$ is the input vector, $\bar{c}$ is the centre (vector) and parameter α (scalar) is a spread constant of RBF function.

There is no general recommendation concerning the input vector size for time series prediction, therefore it has to be done by trial-and-error method - in analyzed case it has to be correlated with a typical time-varying character of the telecommunication traffic shape. Based on extended experiments the following structure of input vector $\bar{x} = [x_1, x_2, x_3, x_4, x_5, x_6, x_7]$ was established:

1. first five coordinates: x_1, x_2, x_3, x_4, x_5 correspond to five values of specified traffic parameter, taken from five time intervals back from the actual time point t: $V(t-5), V(t-4), V(t-3), V(t-2), V(t-1)$, Fig. 4,

2. the x_6 coordinate is the average value $\bar{V}$ of traffic parameter calculated through five time intervals back from the actual time point t,

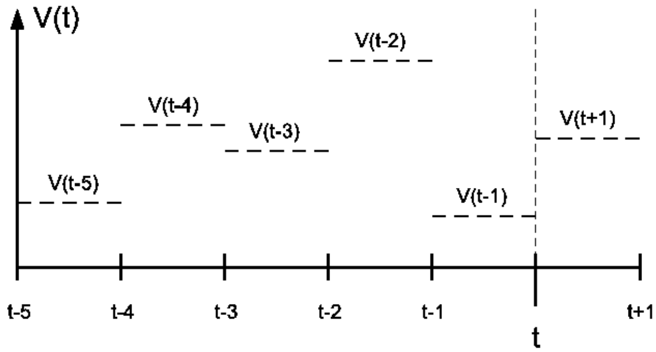

Figure 4. Diagram of time series of telecommunication traffic used for training the Radial Basis Function neural network predictor. Time point (t) represents the moment, when the predicted value for next interval $V(t + 1)$ is calculated

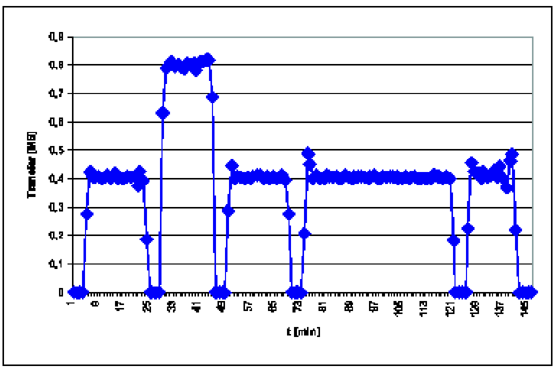

Figure 5. The learning pattern of 5 traffic flows of user activity

3. the x_7 coordinate is the variance σ^2 of traffic parameter calculated through five time intervals back from the actual time point t.

For neural network training example the end user traffic characteristic (average bit rate) has been probed at one minute time interval. This results in a training series as shown in Fig. 5, where long lasting idle intervals have been removed, so the traffic characteristic represents only busy activity of the user.

	X_7 average	X_6 variance	X_5 lag.5	X_4 lag.4	X_3 lag.3	X_2 lag.3	X_1 lag.1
1	$\overline{V_1}$	σ_1^2	V(t-5)	V(t-4)	V(t-3)	V(t-2)	V(t-1)
2	$\overline{V_2}$	σ_2^2	V(t-4)	V(t-3)	V(t-2)	V(t-1)	V(t+1)
3	$\overline{V_3}$	σ_3^2	V(t-3)	V(t-2)	V(t-1)	V(t+1)	V(t+2)
4	$\overline{V_4}$	σ_4^2	V(t-2)	V(t-1)	V(t+1)	V(t+2)	V(t+3)
	...	...	...	...	...	...	...
	...	...	...	...	...	...	...
	...	...	...	...	...	...	...
n-1	V_{n-1}	σ_{n-1}^2	V(t+n-6)	V(t+n-5)	V(t+n-4)	V(t+n-3)	V(t+n-2)
n	V_n	σ_n^2	V(t+n-5)	V(t+n-4)	V(t+n-3)	V(t+n-2)	V(t+n-1)

Figure 6. The input data array for (n) point values prediction. Time point (t) represents the moment, when the prediction of estimated values for (n) points is started.

The training series is used to generate a large number of individual samples, organized in a individual input vector $\bar{x}$, Fig. 6. Each vector consists of five values of traffic parameter, e.g. average bit rate, observed for five time intervals back from the actual time instant t and two additional values, i.e. average and variance of this sequence. After presentation of the input vector $\bar{x}_1$ the predicted value $V'(t+1)$ of the time series for the $t+1$ time interval is calculated by the neural network and the result is compared with the correct learning value $V(t+1)$, which is embedded in coordinates of the next input vector $\bar{x}_2$. This way the learning error is determined and then minimized in the next epoch.

4. Training method

The radial network training process requires solution of three tasks performed in sequence: selection of radial function centres $\bar{c}_n$, determination of radius parameters α_n and weights calculation of the output layer linear neuron. The problem of output layer neurons weights adaptation is made by the so called Green matrix approach exposed to pseudoinversion. The weight vector has the following form: $\bar{w} = G^{+}\bar{d}$, where Green matrix G includes neurons responses to input data in the form of vectors, whose number is equal to the Green matrix number of rows. The $\bar{d}$ vector represents expected output values of prediction system.

The main problem for solution in case of Radial Basis Function neural networks is selection of radial functions parameters, especially c_i centres. This was made using the so-called self-organizing algorithm [12]. According to this algorithm, the learning data set is divided into groups called Voronoi areas, each represented by central point being the mean value of group elements. The centre of each Voronoi area is an equivalent to the radial function centre, so the number of areas is equal to the number of radial functions being controlled by self-organizing algorithm. Initial number of radial functions i.e. hidden layer neurons has direct impact on modelling precision, so bigger input vector size results bigger number of radial functions. The learning procedure performs area centres update after each presentation of one learning data vector $\bar{x}$ [12]. Initial centres selection is performed randomly using uniform distributed numbers. The updating $\bar{c}_i$ centre is realized using the following formula:

$$\bar{c}_i(n+1) = \bar{c}_i(n) + \eta[\bar{x}_n - \bar{c}_i(n)] \tag{2}$$

where the learning factor η decreases its value with elapsed time and takes values much smaller than 1. The other centres are not updated and each vector $\bar{x}$ is presented repeatedly up to final function centres findings.

In Fig. 7 the test result of the neural network proposed for time series prediction (traffic characteristic of the end-user) is shown for two different Radial Basis Functions shapes used: Gaussian-like (the most popular) and thin plate spline RBF.

5. Conclusions

In this paper the possibility of using of artificial neural networks for prediction of TCP/IP packet network user traffic characteristics was presented. It was successfully proved, that proper selection of network type (architecture), training algorithm and training data set, allows to achieve satisfactory results even in case of dynamic, rapidly changing packet traffic source. The Radial Basis Function neural network with self-organizing training algorithm are especially efficient for this purpose, however huge number of flows carried in large packet networks, e.g. Internet, generate performance problems for proposed model in real applications. The number of floating-point operations demand (processor capacity) can be critical issue in case of large systems handling. Additional model for pattern recognition and classification has to be studied for training/testing data set reduction - artificial neural network could be used once again.

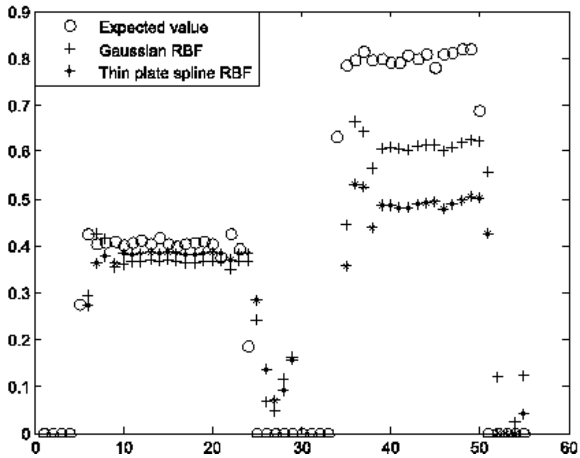

Figure 7. Test results of time series prediction with using the neural network proposed for different RBF shapes used.

References

[1] Cheng, R.-G., Chang, C.-J., Lin, L.-F., *A QoS-Provisioning Neural Fuzzy Connection Admission Controller for Multimedia High-Speed Networks*, IEEE/ACM Transactions on Networking, Vol. 7, No. 1, 111-120, February 1999.

[2] Davoli, F., Maryni, P., *A Two-Level Stochastic Approximation for Admission Control and Bandwidth Allocation*, IEEE Journal on Selected Areas in Communications , Vol. 18, No. 2, February 2000.

[3] Catania, V., Ficili, G., Palazzo, S., and Panno, D., *A Comparative Analysis of Fuzzy Versus Conventional Policing Mechanisms for ATM Networks*, IEEE/ACM Transactions on Networking, Department of Statistics, Vol. 4, No. 3, June 1996.

[4] Bivens, J. A., Embrechts, M. J., and Szymański, B. K., *Network Congestion Arbitration and Source Problem Prediction Using Neural Network*, Smart Engineering System Design, Vol. 4, pp. 243–252, 2002.

[5] Kopertowski, Z., Burakowski, W., and Bąk, A., *Artificial neural networks application in telecommunication*, Telecommunication review, Vol. 4 1994, pp. 191–197 (in Polish).

[6] Papir, Z., *Telecommunication traffic and packed networks overload*, WKŁ, Warsaw, 2001 (in Polish).

[7] Grzech, A., *Traffic management in telecommunication networks*, WPW, Wrocław, 2002 (in Polish).

[8] Stevard, R., Xie, Q., Morneault, K., Sharp, C., Shwarzbauer, H., Taylor, T., Rytina, I., Kalla, M., Shang. L., and Paxton, V., *Stream Control Transmission Protocol*, RFC 2960, October 2000.

[9] Box, G., Jenkins, G. M., and Reinsel, G., *Time Series Analysis: Forecasting and Control*, Prentice Hall, February 1994.

[10] Zhao, G.-F., Tang, H., Xu, W.-B., and Zhang, Y.-H., *Application of Neural Networks for Traffic Forecasting in Telecom Networks*, Proc. of Third Int. Conference on Machine Learning and Cybernetics, Vol. 4, pp. 2607-2611, Shanghai, 26–29 August 2004.

[11] Zaleski, A. and Kacprzak, T., *Radial based neural network for traffic prediction in telecommunication networks*, Proc. of 16th Polish Teletraffic Symposium, pp. 153-156, Łódź, 24–25 September 2009.

[12] Osowski, S., *Artificial neural networks for information processing*, OWPW, Warsaw, 2000 (in Polish).