Qian GAO, Guangxia LI
Institute of Communication Engineering, PLA Univ. of Sci. & Tech.

A Traffic Prediction Method based on ANN and Adaptive Template Matching

Abstract. *Considering the significant burstiness in self-similar network traffic, this paper presents a one-step traffic prediction method based on ANN (artificial neural network) and adaptive template matching. Firstly control chart is used to decompose the original traffic into non-burst part and burst part. Then the two parts are predicted by ANN and adaptive template matching method, respectively. Finally, the prediction of the original traffic is obtained by fusing the predictions of the two parts. Simulation results show that the accuracy of method proposed in this paper is much higher compared with the traditional traffic prediction methods.*

Streszczenie. *Artykuł prezentuje metodę przewidywania przepływu informacji bazującą na sztucznych sieciach neuronowych i adaptacyjnym dopasowaniu szablonu. Początkowo przepływ dzielony jest na pogrupowany i nie pogrupowany. Następnie analizuje się obie części. Wreszcie dokonuje się fuzji rezultatów. Rezultaty symulacji wskazują że proponowana metoda ma lepszą dokładność niż metody tradycyjne. (**Metoda przewidywania przepływu informacji bazująca na sztucznych sieciach neuronowych i adaptacyjnym dopasowaniu szablonu**)*

Keywords: Traffic prediction; Control chart; Artificial neural network (ANN); Adaptive template matching
Słowa kluczowe: przewidywanie przepływu informacji, systemy komunikacyjne, sztuczne sieci nauronowe.

1 Introduction

In network management, traffic prediction plays important roles in resource allocation, congestion control and so on. Especially in broadband satellite communications system, access control and bandwidth allocation are centralized decision-making by the NCC (Network Control Center). In face of larger network delay, the decisions made by NCC are based on the traffic prediction results for each user at the next time [1]. Thus, the accuracy of traffic prediction will directly affect the performance of the entire network.

With the development of broadband technology, multimedia service has become the most important type of service. Different from the traditional telephone service characteristic, multimedia service shows new feature: self-similarity. One of the most prominent characteristics of self-similarity is that traffic presents obvious burstiness at all time scales. In addition, for the traditional telephone traffic which obeys Poisson distribution, traffic burstiness is relatively obvious in smaller time scale. And as the observation interval getting bigger, traffic burstiness will reduce significantly [2].

There has been a variety of prediction models proposed until now. These models can be roughly divided into two categories:

- Time series models. There are Markov model [3, 4], ARMA (Auto Regressive Moving Average) model [5, 6], ARIMA (Auto Regressive Integrated Moving Average) model [7], FARIMA (Fractional Auto Regressive Integrated Moving Average) model [8] and so on.
- Prediction model based on neural network or fuzzy neural network [9-11].

In addition to the above models, for the sake of solving the long-range dependency characteristics existing in the self-similar network traffic, the signal analysis methods based on DWT (Discrete Wavelet Transform) [12] have been used for network traffic prediction.

The methods mentioned above have their merits for improving the accuracy of traffic prediction, but the results are still satisfactory. The reasons why it is difficult to predict accurately can be attributed to the following two problems:

(1) There is randomicity inherited in traffic data which means it will not obey certain rules exactly. Then to some extent, it is inevitable for the emergence of prediction error;

(2) There are many bursts in network, and traffic will increase sharply when burst occurs. It is difficult to take into account simultaneously both the burst traffic and non-burst traffic within a single prediction model.

For problem (1), we can choose prediction model flexibly which is suitable for specific traffic data; for problem (2), this paper presents a method combining adaptive artificial neural network with template matching. More specifically, the burstiness of traffic data will be detected firstly. Then, the original data will be decomposed into burst data traffic and non-burst data and be predicted, respectively. Finally, the prediction of the original traffic is obtained by fusing the predictions of the two parts. The detection of burst data is achieved by control chart. And adaptive template matching method will be used for the predicting the burst data which is relatively difficult for traditional prediction model due to its sudden sharp changes. While the prediction of non-burst data is performed by ANN (Artificial Neural Network) which takes powerful non-linear modeling ability. Specific methods and simulation results will be described in subsequent sections in detail.

The structure of this paper is as follows: Section 2 introduces the basic knowledge of control chart and data decomposed method briefly; Section 3 describes the proposed traffic prediction method based on ANN and adaptive template matching; Section 4 validates the accuracy of proposed method through simulation based on real traffic data; the conclusion will be given in Section 5.

2 Control chart-based burst data decomposition

2.1 Control chart

Control chart is a graphical tool which is used to indicate the current random process whether is in controllable state. It is an important statistical quality control method and tool [13]. The stability state can be interpreted as a state in which a process has displayed a certain degree of consistency in the past and is expected to continue to be seen in the future. In the control field, when a process is in controllable state, generally it doesn't need to modify the process control parameters. However, when a process is in uncontrollable state, if we use current data to predict future data, the accuracy of the prediction can not be guaranteed.

One of the implicit assumptions in control chart is that "An event of very small probability will not happen on a particular trial". Supposed that a random process is stationary Gaussian process, whose mean is μ , variance is σ^2 , the random variation at t time point is x_t , so:

(1) $$P\left(\left|\frac{x_t-\mu}{\sigma}\right|<3\sigma\right)=99.73\%$$

It means that the samples are in interval $(-3\sigma,3\sigma)$ with the probability of 99.73%. While the probability for sample falling outside this interval is just 0.27%. Therefore, the samples out of this interval can be considered as "event of very small probability", which should not occur in controllable state.

According to this theory, three baselines of control chart can be determined by the mean and variance of the random process: CL (Central Line), UCL (Upper Control Limit) and LCL (Lower Control Limit). CL can be set as mean, UCL and LCL can be set 3σ above and below the mean, respectively. When observation data is obtained, these three baselines can constitute constraint area for normal fluctuations, as shown in Fig. 1. The samples in this interval are regarded as in controllable state, otherwise, as in uncontrollable state.

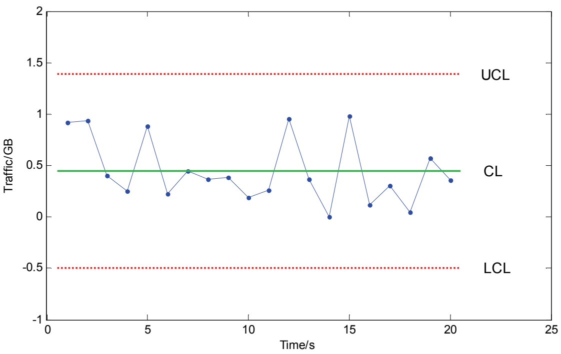


Fig.1.Control chart

In practical used control chart, the determination of three baselines is much more complicated. μ and σ can be calculated by different methods, while the 3σ theory which is used to decide UCL and LCL is remain unchanged, which means that UCL and LCL are set as 3σ above and below μ, respectively. The commonly used control chart includes:

1) X-Bar and R Chart

When multiple measurements of random process are obtained, this type of control chart can be adopted to check for violations. The measurements are presented in Equation (2), where x_{ij} represents the j th time sample of the process in the i th measurement. There are m measurements of the process, and each measurement contains n samples.

(2) $$\begin{bmatrix} x_{11},x_{12},\cdots,x_{1j},\cdots,x_{1n} \\ x_{21},x_{22},\cdots,x_{2j},\cdots,x_{2n} \\ \vdots \quad \vdots \quad \vdots \\ x_{i1},x_{i2},\cdots,x_{ij},\cdots,x_{in} \\ \vdots \quad \vdots \quad \vdots \\ x_{m1},x_{m2},\cdots,x_{mj},\cdots,x_{mn} \end{bmatrix}$$

Here, the control chart is performed based on the means of all the measurements, namely,

(3) $$\bar{x}_j=\frac{1}{m}\sum_{i=1}^{m}x_{ij}$$

Thus, CL is calculated by Equation (4):

(4) $$\text{CL}=\frac{1}{n}\sum_{j=1}^{n}\bar{x}_j$$

UCL and LCL are obtained using Equation (5) and (6):

(5) $$\text{UCL}=\text{CL}+3\sigma_{XR}=\text{CL}+\frac{3}{n}\sum_{j=1}^{n}r_j$$

(6) $$\text{LCL}=\text{CL}\text{-}3\sigma_{XR}=\text{CL}-\frac{3}{n}\sum_{j=1}^{n}r_j$$

where,

(7) $$r_j=\left|r_{j\max}-r_{j\min}\right|$$

$r_{j\max}$ and $r_{j\min}$ are the maximum and minimum value of the m measurements samples at time point j, respectively, where " $|\bullet|$ " means calculating the absolute value.

2) Individual X and Moving Range Chart

When there is only one measurement, Individual X and Moving Range Chart can be used. Here, CL is mean of samples in the measurement, as shown Equation (8):

(8) $$\text{CL}=\frac{1}{n}\sum_{j=1}^{n}x_j$$

UCL and LCL are obtained by Equation (9) and (10):

(9) $$\text{UCL}=\text{CL}+3\sigma_{IRm}=\text{CL}+\frac{3}{n'}\sum_{j=1}^{n'}r_j^M$$

(10) $$\text{LCL}=\text{CL}-3\sigma_{IRm}=\text{CL}-\frac{3}{n'}\sum_{j=1}^{n'}r_j^M$$

where

(11) $$r_j^M=\left|r_{j\max}^M-r_{j\min}^M\right|$$

$r_{j\max}^M$ and $r_{j\min}^M$ are the maximum and minimum value among samples $x_j\sim x_{j+T}$ respectively, where T is nonnegative integer, and $n'=n-T$.

2.2 Control chart-based burst traffic decomposition method

Fig. 2 shows a segment of network traffic data, which is taken from "LBL-tcp-3. tcp" (which can be obtained from: http://ita.ee.lbl.gov/html/contrib/LBL-TCP-html). "LBL-tcp-3. Tcp" records traffic data for 2 hours.

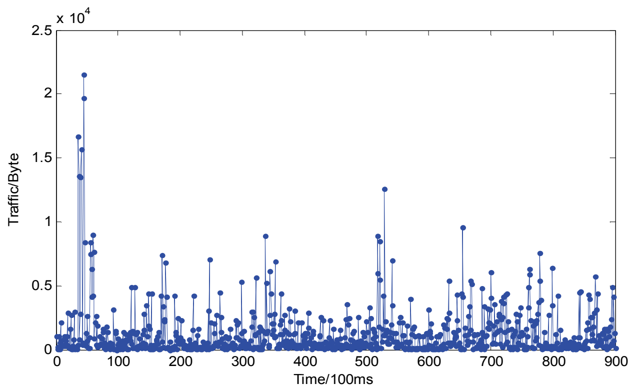


Fig. 2. "LBL-tcp-3. tcp" 90s data segment with 100ms interval

Fig. 2 shows 90s of this trace, in which the traffic values of every 100ms are presented. From Fig. 2, it can be known that this traffic is very bursty.

In the view of process control, these bursts can be considered as uncontrollable data. Therefore, we can employ control chart to identify and decompose them from the original traffic. As there is only one measurement of the traffic, "Individual X and Moving Range Chart" method is adopted for this task. When slide window T is set to 5, the burst identification results obtained by control chart are shown in Fig. 3. The points which are marked by red circles fluctuate out of controllable area. These points are treated as burst traffic.

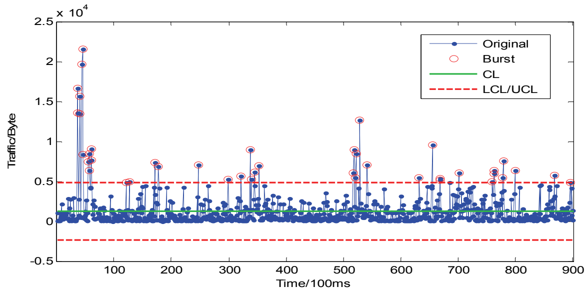


Fig.3. Determination of burst through control chart

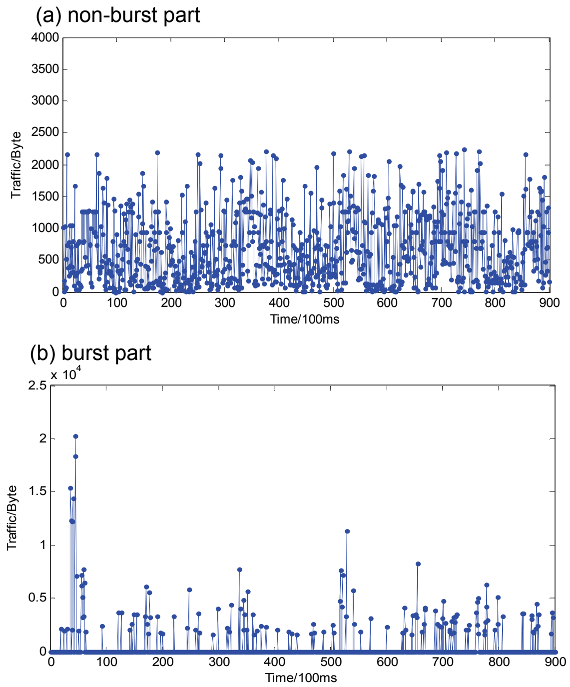


Fig.4. Decomposition traffic

In this paper, hard-limiting process is employed to wipe out the burst traffic. More specifically when x_j is the burst point identified by control chart, it will be dealt with in the way as shown in Equation (12):

(12) $$x_j = \mathrm{CL}$$

In order to guarantee all the burst data is removed, we employ the control chart to the original traffic in an iterative manner for some times. Then the non-burst traffic is achieved as shown in Fig. 4 (a). Meanwhile the burst part is also decomposed from the original traffic which is presented in Fig. 4 (b).

As shown in Fig. 4, it is obvious that the non-burst traffic is more stable than the original traffic, and the burst part is sparsely located in the time axis. According to these characteristics of decomposed results, in next section, we will propose a novel prediction method based on artificial neural network and adaptive template matching.

3 ANN and adaptive template matching-based traffic prediction method

By control chart method, the non-burst traffic and burst traffic is extracted from the original traffic trace. However due to the hard-limiting process, non-linearity is introduced into the non-burst traffic data. So, this paper adopts ANN to predict these data. On the other side, it appears that the traditional correlation-based prediction models are not suitable for the prediction of the burst data. Thereby, this paper proposes a template matching-based prediction method for burst data prediction.

3.1 ANN-based traffic prediction

This scheme uses multilayer feedforward neural network to modeling non-burst data. Compared with other linear models such as ARMA, ANN has the ability to model the non-linear relationship in the traffic data.

As ANN with three layers can fit any function, so we use a three-layer ANN for the prediction task. The structure of ANN can be expressed as " $a-bN-cL$ ", where a is number of inputs, b is the number of hidden neurons, c is the number of output neurons. And N indicates that the neurons of this layer adopt non-linear " tansig " excitation function, L indicates that the neurons adopt linear excitation function.

For the ANN training, BP (Back Propagation) algorithm is adopted to get the neural network weights with minimum output error through multiple training iterations.

3.2 Adaptive template matching-based traffic prediction

Through analysis, it can be seen that the decomposed burst data are zero at most of the time points, and the non-zero bursts are dispersed in the entire time span. Therefore, in order to perform traffic prediction, burst traffic should be segmented to obtain the basic unit: traffic cluster. In the burst data, the traffic which is composed of non-zero bursts at continuous time points is defined as the cluster. In this way, the non-zero burst with zero traffic in adjacent time points also forms an independent cluster.

However, due to the short duration of the clusters, prediction can hardly be performed based only on the burst traffic clusters themselves. So the original traffic data of three time points before the cluster and one time point after the cluster are added to cluster to help the prediction. The additional traffic data and cluster together constitute the prediction traffic cluster.

Then the prediction method based on adaptive template matching can be summarized as follows:

1) Building prediction traffic cluster template library based on prediction cluster in training data;

Group the prediction cluster contained in training data according to the number of time points they contained, and arrange prediction traffic cluster in time sequence in each group. Consequently, obtain the following template library:

(13) $$\mathbf{P} = \begin{Bmatrix} \mathbf{bc}_{51}, \mathbf{bc}_{52}, \cdots, \mathbf{bc}_{5j}, \cdots, \mathbf{bc}_{5N_5} \\ \mathbf{bc}_{61}, \mathbf{bc}_{62}, \cdots, \mathbf{bc}_{6j}, \cdots, \mathbf{bc}_{6N_6} \\ \vdots \qquad \vdots \qquad \vdots \\ \mathbf{bc}_{i1}, \mathbf{bc}_{i2}, \cdots, \mathbf{bc}_{ij}, \cdots, \mathbf{bc}_{iN_i} \\ \vdots \qquad \vdots \qquad \vdots \\ \mathbf{bc}_{M1}, \mathbf{bc}_{M2}, \cdots, \mathbf{bc}_{Mj}, \cdots, \mathbf{bc}_{MN_M} \end{Bmatrix}$$

where, $\mathbf{bc}_{ij}$ is the j th prediction cluster in training data who contains i time points, N_i is the number of prediction clusters who contains i time points in training data, M is maximum length of prediction cluster in training data. Since that each traffic cluster contains 4 non-burst traffic data in addition to the burst, the length of the prediction cluster is 5 time points at least.

2) In the predicting period, determine the burst by control chart and then choose the prediction method accordingly for the burst traffic and non-burst traffic;

Supposed that the traffic value at time point n is y_n. If

$y_n > \mathrm{UCL}$, it means that the current traffic is burst, then the template matching method will be adopted to predict traffic at the next time point. Conversely, if $y_n < \mathrm{UCL}$, the ANN method will be adopted for prediction.

3) Using adaptive template matching method to predict burst traffic;

If there are T_n^b continuous time points before y_n where the traffic data are burst, then a vector y_n^b is built with these bursts and the three original traffic data before the burst. Then length of y_n^b is $T_n^b + 4$. Vector y_n^b is shown by Equation (14):

(14) $$\mathbf{y}_n^b = \left[y_{n-T_n^b-3}, y_{n-T_n^b-2}, y_{n-T_n^b-1}, y_{n-T_n^b}, y_{n-T_n^b+1}, \cdots, y_n \right]$$

where T_n^b is a nonnegative integer.

The next step is to match y_n^b adaptively with prediction traffic cluster template elements in $\mathbf{P}$ whose length is longer than $T_n^b + 4$. Through matching process, the template element with minimum matching error is picked out. Finally the prediction of the burst in next time point is obtained based on the value of template in corresponding time point. The matching process is performed as shown in Equation (15):

(15) $$\min_{i,j} \left\| \mathbf{y}_n^b - \lambda \mathbf{bc}'_{ij} \right\|_2$$

in (15) λ is defined by Equation (16):

(16) $$\min_{\lambda} \left\| \mathbf{y}_n^b - \lambda \mathbf{bc}'_{ij} \right\|_2$$

where $\mathbf{bc}'_{ij}$ is the former $T_n^b + 4$ time points traffic value of $\mathbf{bc}_{ij}$. λ is traffic scale transformation factor. By adjusting λ , y_n^b could match with $\mathbf{bc}'_{ij}$ adaptively. After finding the optimal prediction traffic cluster template, the prediction can be obtained by multiplying the traffic value of the optimal template at $T_n^b + 5$ th time point with λ .

As mentioned above, the adaptive template matching-based method performs prediction making use of similarity between traffic bursts. Generally speaking, it can be assumed that the closer two bursts are in time domain, the more related they should be. Thus, during template searching, add restrictions that search only three templates which are closest to time point n in each row of template library $\mathbf{P}$.

4) Updating the prediction traffic cluster template library $\mathbf{P}$;

If the traffic at time point n has meet the formation conditions of prediction traffic cluster, then a new prediction traffic cluster template will be created and added to the template library $\mathbf{P}$ at appropriate location;

5) Return to step 2) for the next time point traffic prediction.

4 Experimental results

The one-step prediction experiments are carried out based on a segment of "LBL-tcp-3. tcp". 90s is used from that trace and the traffic observation interval is set to 100ms. So there are 900 traffic data used in our experiments, as shown in Fig. 2. In experiments, the former 800 traffic data are employed to build ANN model as well as original prediction traffic cluster template library. And the latter 100 time points are used for predicting.

In order to evaluate prediction performance, the MSE (Mean Squared Error) is used as evaluating method in this paper, which is:

(17) $$e_{mse} = \frac{1}{K} \sum_{k=1}^{K} \left(y_k - \hat{y}_k \right)^2$$

Where y_k is the actual traffic value at the k th time point, $\hat{y}_k$ is the one-step prediction result, K is the total prediction number.

In the experiment, the ANN neurons number of each layer is set to 8-45N-1L which is obtained through several trials. For comparison purpose, ANN with structure 8-60N-1L, ARMA (4,2) and ARIMA (1,1,3) methods are also carried out to perform the one-step prediction based the same training and testing data. For the ARMA and ARIMA model, their orders are selected by utilizing AIC (Akaike Information Criterion) principle.

On account of the traffic data is nonnegative, the prediction results are set to 10 if they turn out to be negative.

The prediction results of the four prediction methods are shown in Table 1. It can be seen, the proposed method outperforms other methods at prediction accuracy.

Table 1. Performance comparison of four traffic prediction method

	MSE
the proposed method	1.603×10^6
ARMA (4,2)	3.712×10^6
ARIMA (1,1,3)	3.311×10^6
ANN (8-60N-1L)	1.735×10^6

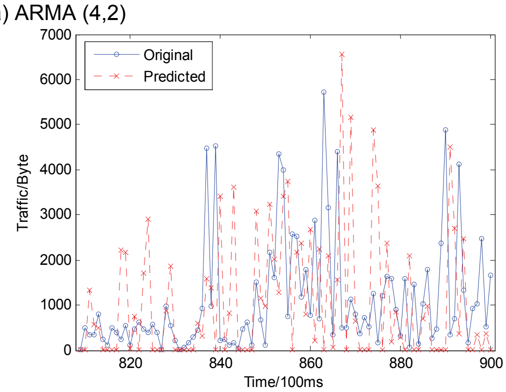
The prediction results of ARMA (4,2), ARIMA (1,1,3) and ANN (8-60N-1L) are shown as Fig. 5. Fig. 6 shows the prediction results of our proposed method.

5 Conclusions and analysis

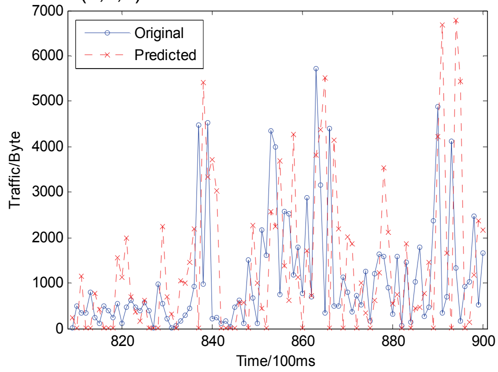
This paper presents a traffic prediction method based on ANN and adaptive template matching method. Firstly, the original traffic is decomposed into burst traffic and non-burst traffic by control chart. Then these two parts are predicted using ANN and adaptive template matching method respectively. Finally the prediction results of the original traffic are obtained by combining the prediction of the two parts.In the proposed method, the burst traffic has been successfully decomposed from the original traffic by control chart. In this way, the non-burst traffic can be predicted more accurately without the impact of burst data. Besides the burst traffic can also be predicted taking full consideration of its sudden sharp changes by adaptive template matching method.

Be noted that, in this method, no special consideration has been given to the nonstationarity of traffic data. But when the traffic is nonstationary, such as having obvious trend, differential processing can be introduced to alleviate this nonstationarity.

(a) ARMA (4,2)



(b) ARIMA (1,1,3)



(c) ANN (8-60N-1L)

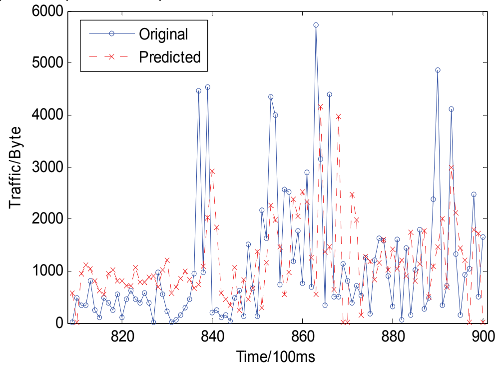
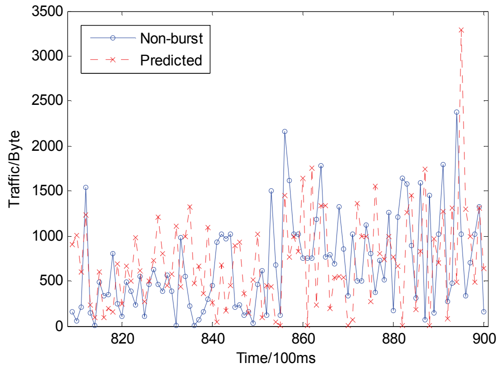


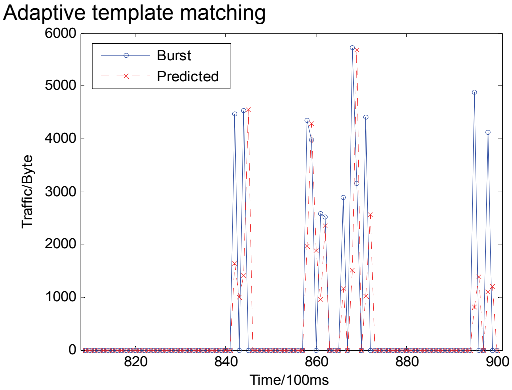
Fig.5. Prediction results of three methods

In addition, as the prediction process of burst data requires determining the burstiness of traffic in the first place, it inevitably causes some inaccuracy for predicting the burst cluster at the first time point. We will search for more effective method to solve this problem in the future study.

(a) ANN



(b) Adaptive template matching



(c) The combined two methods

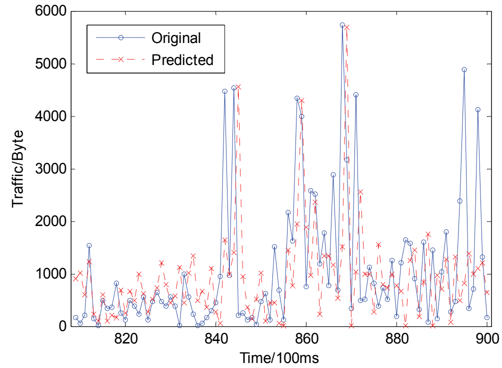


Fig.6. Prediction results of the proposed method

Acknowledgement

The authors wish to acknowledge the National Natural Science Foundation of China (60972061 and 61032004), Natural Science Foundation of Jiangsu Province (BK2009060) and "Triple Three" High Level Talent Development Plan of Jiangsu Province.

REFERENCES

[1] Delli Priscoli F, Pompili D., A demand-assignment algorithm based on a Markov modulated chain prediction model for satellite bandwidth allocation, Wirel. Netw., (2009), No.15, 999-1012

[2] Leland W E, Taqqu M S, Willinger W, et al., On the self-similar nature of Ethernet traffic (extended version), Networking, IEEE/ACM Transactions on, (1994), No. 2(1): 1-15

[3] Yoshihara T, Kasahara S, Takahashi Y., Practical Time-Scale Fitting of Self-Similar Traffic with Markov-Modulated Poisson Process, Springer Netherlands, (2001), No. 17, 185-211

[4] Nogueira A, Salvador P, Valadas R, et al., Markovian Modelling of Internet Traffic: Lecture Notes in Computer Science, Kouvatsos D. Springer Berlin / Heidelberg, (2011), No. 5233, 98-124

[5] I. S. Gradshteyn, Discrete Time Stochastic Systems-Esitmation and Control, Englewood Cliffs, NJ: Prentice-Hall, (1994)

[6] Shih Y C, Hsiao-Chun W., Novel Fast Computation Algorithm of the Second-Order Statistics for Autoregressive Moving-Average Processes, Signal Processing, IEEE Transactions on, (2009), No. 57(2), 526-535

[7] Man-Chun T, Wong S C, Jian-Min X, et al., An Aggregation Approach to Short-Term Traffic Flow Prediction, Intelligent Transportation Systems, IEEE Transactions on, (2009), No. 10(1), 60-69

[8] Yantai S, Zhigang J, Lianfang Z, et al., Traffic prediction using FARIMA models, ICC '99, IEEE International Conference on, (1999)

[9] Lin X, Xiao-Hu G, Chuang L, et al., A New Hybrid Network Traffic Prediction Method, GLOBECOM 2010, 2010 IEEE Global Telecommunications Conference, (2010)

[10] LI R, CHEN J, LIU Y, et al., WPANFIS: combine fuzzy neural network with multiresolution for network traffic prediction, The Journal of China Universities of Posts and Telecommunications, (2010), No. 17(4), 88-93

[11] V. A, J. A B., Multiresolution FIR neural-network-based learning algorithm applied to network traffic prediction, Systems, Man, and Cybernetics, IEEE Transactions on, (2006), No. 36(2), 208-220

[12] Wei Y, Wang J, Wang C, Network Traffic Prediction Based on Wavelet Transform and Season ARIMA Model, Lecture Notes in Computer Science, Liu D, Zhang H, Polycarpou M, et al. Springer Berlin / Heidelberg, (2011), No. 6677, 152-159

[13] John S Oakland, Statistical Process Control, Third Edition: A Really Practical Guide, Butterworth-Heinemann, (1996)

***Authors**:*
Qian GAO, dr, Military Satellite Communications LAB, Institute of Communications Engineering, Biaoyin 2, Yudao Street, Nanjing, China, 210007, Email: gaoxiongmao1234@163.com;
Guangxia LI, prof., Military Satellite Communications LAB, Institute of Communications Engineering, Biaoyin 2, Yudao Street, Nanjing, China, 210007.