

A New Hybrid Network Traffic Prediction Method

Lin Xiang¹, Xiao-Hu Ge¹, Chuang Liu¹, Lei Shu², Cheng-Xiang Wang³

¹Dept. Electronics & Information Engineering, Huazhong University of Science & Technology, Wuhan, P.R.China.

²Nishio Lab., Dept. Multimedia Engineering, Graduate School of Information Science & Technology, Osaka University, Japan.

³JRI-SIP, School of Engineering & Physical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, U.K.

¹xhge@mail.hust.edu.cn, ²lei.shu@ist.osaka-u.ac.jp, ³cheng-xiang.wang@hw.ac.uk

Abstract—How to predict the self-similar network traffic with high burstiness is a great challenge for network management. The covariation orthogonal prediction could effectively capture the burstiness in the network traffic, and the artificial neural network prediction could adapt the network traffic change by self-learning. To improve the prediction accuracy, we propose a new hybrid network traffic prediction method based on the combination of the covariation orthogonal prediction and the artificial neural network prediction. Through empirical study, the accuracy of the new prediction method can be effectively improved seen from the mean and the prediction error.

Index Terms—Traffic Prediction; Covariation Orthogonal (CO); Artificial Neural Networks (ANNs); Self-similarity; Burstiness;

I. INTRODUCTION

A number of studies have demonstrated that network traffic is self-similar in nature, showing high burstiness in a wide range of time scales, and obeys in the heavy-tail distribution [1]. Similarly, Errors encountered in wireless communication systems are bursty in nature and the burst error statistics have been well studied in [2],[3],[4]. The alpha-stable model can capture the self-similarity and heavy-tailness of network traffic simultaneously [5], and has played an increasingly important role in traffic prediction.

However, the infinite variance of self-similar traffic process, known as the high variability, will devaluate the classical least mean square error criterion for alpha-stable model based traffic predictors. To cope with this challenge, the minimum dispersion (MD) criterion and the covariation orthogonal (CO) criterion [6],[7],[8],[9],[10],[11] have been applied instead. These prediction schemes manage to explore the correlations between alpha-stable traffic processes with the dispersion or covariation. Then, the possible relationship between the future traffic process and the past traffic process can be developed using estimation techniques, which can be applied for forecasting future network traffic. The CO linear prediction [8] can be effective in predicting the bursty changes of the self-similar network traffic. However, the mean of the predicted traffic under this scheme often goes below the actual level, resulting the limited prediction accuracy.

Different from the alpha-stable method, the artificial neural networks (ANNs) can “self-learn” the traffic characteristics by following a set of learning rules to adapt its weight coefficients and make predictions without any prior knowledge of the characteristics of network traffic. Different forms of

ANNs have been applied for real traffic prediction. However, there is no single ANN form that can capture all the traffic characteristics [12].

In this paper, our work is to combine the CO linear prediction scheme with the ANN approach so as to raise the prediction accuracy of the CO approach. As mentioned above, the CO linear prediction can capture the bursty changes of network traffic while under-predict the future traffic values; and the ANN approaches are capable of predicting self-similar traffic with high precision while need combination with other methods. Considering the pits and falls of both approaches, we propose a new hybrid network traffic prediction method to improve the CO prediction method and to raise the prediction accuracy. The main contributions of this paper are as follows:

- 1) A simple hybrid prediction scheme based on the CO criterion and the ANNs is proposed for predicting self-similar network traffic.
- 2) Performance of the proposed method is evaluated through empirical study using real traffic traces, in terms of the self-similar and bursty characteristics of the predicted traffic and the prediction error.
- 3) The effects of both parts in the hybrid prediction are analyzed, including their limitations in predicting bursty network traffic.

The rest of this paper is organized as follows: Section II introduces selected work on traffic prediction; Section III describes the system model and principles of each prediction part; Section IV empirically studies the prediction performances with real traffic traces collected in wireless local area networks; finally, Section V concludes the paper.

II. RELATED WORK

The research on alpha-stable model based prediction has mainly focused on the linear prediction form [6],[7],[8],[9],[10],[11]. However, the computation of the prediction coefficients is not straightforward since the second order moments of alpha-stable process do not exist. Based on the MD criterion, Gallardo worked out a novel method to compute the coefficients [6] and usually provide a robust performance suggested in [9]. However, this method results in a biased forecast with non-unique closed-form prediction results. Therefore the minimization of dispersion is not optimal [11]. To solve this problem, an optimizing approach using steepest decedent method has been proposed to search the actual MD

point at a faster speed [10] and applied in dynamical resource allocations [11]. Another work [7] considers the compound prediction of wireless network traffic using the MD criterion. Working from a new direction, Ge proposed the unbiased CO linear prediction method in [8]. However, as previously mentioned, the prediction accuracy of this method needs to be improved to adapt the busy traffic. This is the important motivation of our work in this paper.

The ANN method predicts network traffic by performing non-linear mappings between the future traffic values and the past and present traffic values. However, no single ANN form can capture all the characteristics of network traffic due to its complexity. Current research has been devoted to develop more powerful ANN forms, such as the novel BP networks [13], or to combine different forms of ANNs for self-similar traffic prediction [14].

In this paper, the CO criterion and the non-linear ANN prediction approach are combined to improve the CO linear prediction method. The CO prediction can capture the bursty changes while the ANN can gain high prediction accuracy. The hybrid design aims to raise the prediction accuracy by making the trade-offs between these two methods.

III. SYSTEM MODEL

Network traffic prediction is the process of mapping past (and present) traffic values to future traffic values through linear or non-linear mapping functions as shown in Eq.(1),

$$\hat{X}(t+k) = F[X(t), X(t-1), \dots, X(t-p+1)] \quad (1)$$

where the function F maps the past p traffic values $X(t), X(t-1), \dots, X(t-p+1)$ into the k -step-ahead traffic value $\hat{X}(t+k)$. The design of traffic prediction scheme mainly concerns in constructing or devising the proper mapping functions.

The simple serial combination of the CO linear prediction and the ANN prediction methods is considered here and the system model is shown in Fig. 1. The input of the CO part is the normalized time series X_i of network traffic; after the CO linear prediction, its output is the time series X_c , which is then the input for the ANN prediction; and the final prediction result is the output time series X_o . The mathematical model for the system can be simply represented as:

$$X_c = F_1(X_i) \quad (2)$$

$$X_o = F_2(X_c) = F_2(F_1(X_i)) \quad (3)$$

where F_1 and F_2 are prediction functions, representing the CO prediction and the ANN prediction respectively. The principles of both prediction parts are covered in the following parts.

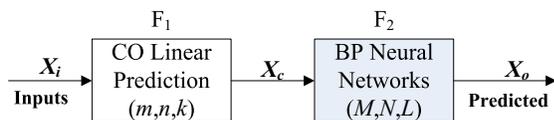


Fig. 1. System Model.

A. The CO Linear Prediction

The covariance function is a powerful tool in the study of Gaussian random elements, but it is not defined in the alpha-stable processes when $\alpha < 2$. Under this condition, the covariation is designed to replace the covariance and its definition is as follows:

Let X and Y be stable random vectors with $1 < \alpha < 2$ and spectral measure Γ , and (μ, ν) denotes the polar coordinates of the joint process (X, Y) on the \mathbb{R}^2 unit circle S_2 , the covariation of X on Y is denoted as:

$$[X, Y]_\alpha = \int_{S_2} \mu \nu^{<\alpha-1>} \Gamma(d\mu, d\nu) \quad (4)$$

When $[X, Y]_\alpha = 0$, the vector X is covariation orthogonal to the vector Y .

Suppose X_1, X_2, \dots, X_n are the n sample values of the input traffic series X_i , the coefficient set is a_1, a_2, \dots, a_m are the n . Reference [8] works out the k -step-ahead linear prediction scheme:

$$F_1 : \hat{X}_{n+k} = \sum_{i=1}^m a_i X_{n+1-i} \quad (5)$$

where \hat{X}_{n+k} is the predicted values, and the coefficients are calculated according to Eq. (6). This approach is noted as a (m, n, k) CO linear predictor, meaning a k -step-ahead linear prediction with m coefficients and n sample values.

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n X_i X_i^{<p-1>} & \sum_{i=2}^n X_{i-1} X_i^{<p-1>} & \dots & \sum_{i=m}^n X_{i+1-m} X_i^{<p-1>} \\ \sum_{i=2}^n X_i X_{i-1}^{<p-1>} & \sum_{i=2}^{n+1} X_{i-1} X_{i-1}^{<p-1>} & \dots & \sum_{i=m}^{n+1} X_{i+1-m} X_{i-1}^{<p-1>} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=m}^n X_i X_{i+1-m}^{<p-1>} & \sum_{i=m}^{n+1} X_{i-1} X_{i+1-m}^{<p-1>} & \dots & \sum_{i=m}^{n+m+1} X_{i+1-m} X_{i+1-m}^{<p-1>} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=k+1}^n X_i X_{i-k}^{<p-1>} \\ \sum_{i=k+2}^n X_i X_{i-k-1}^{<p-1>} \\ \vdots \\ \sum_{i=k+m}^n X_i X_{i+k-m}^{<p-1>} \end{bmatrix} \quad (6)$$

B. The ANN Prediction

A (M, N, L) 3-layer common back-propagation (BP) neural networks are chosen for traffic prediction, where M, N, L represents the number of neurons at each layer. Different from the CO approach, the prediction function F_2 of the ANN approach is not explicitly computed, but "self-learned". Specifically, each layer follows the specific learning rules and adjusts the weights until the mean square error between the expected output and the actual output becomes lower than a pre-determined value. Then, the BP neural network can be used for traffic prediction. In the system model, fractional samples of the time series X_c are used for training the BP neural network.

IV. THE PROPOSED HYBRID SCHEME

Based on the system model, the detailed hybrid prediction process is designed as follows, divided into two periods:

Prediction setup: in this period, the system is initialized and then the BP neural networks are trained to fit the real traffic until the pre-determined accuracy is achieved.

Step 1: Decide the architecture and parameters for the BP neural networks, such as the triplet (M, N, L) representing the number of nodes at each layer and the mean square error between the expected and the actual output, i.e., the pre-defined prediction accuracy. Besides, decide the parameter triplet (m, n, k) for the CO prediction part, where m is the number of the prediction coefficients, n is the norm of the sample space $\{X_1, X_2, \dots, X_n\}$ and k is the prediction step.

Step 2: Initialize the weights randomly.

Step 3: Train the BP neural network and stop training until the pre-defined accuracy is achieved.

Prediction: in this period, the well-trained BP neural networks can be used for traffic prediction. This process can be further divided into two stages—the CO prediction stage and the ANN prediction stage, with each containing several steps.

Step 4: The first $1 - n$ samples are picked up from the input traffic series X_i to calculate the prediction coefficients according to Equ. (6).

Step 5: The k -step-ahead prediction result are calculated under the CO criterion by Equ. (5).

Step 6: The $2 - (n + 1)$ samples are picked up for another prediction. Go on to Step 5 and this process continues. This is the first stage of the hybrid prediction and the prediction results are the time series X_c .

Step 7: The M samples of the time series X_c are input to the BP neural networks for the second stage prediction. Then shift to another M samples and continue the ANN prediction. The final prediction results are the time series X_o .

By combining these two methods, our proposed scheme aims to raise the prediction ability. Specifically, the CO prediction part is used as a coarse prediction and aims to capture the variation trend of the self-similar traffic; while the NN part is used as a fine prediction so as to raise the prediction accuracy. The prediction accuracy is evaluated in the following section.

V. SIMULATION EXPERIMENTATION

In this part, the prediction performance of the hybrid scheme is evaluated by two traffic traces and compared with the CO linear prediction scheme. The traces are collected from a real wireless local area network (WLAN) deployed during the 62nd IETF Plenary Session [15] on March 10, 2005. The previous study has revealed the self-similarity of the collected traffic [16]. Both traces, with different degrees of fluctuations or burstiness, are aggregated at 0.1s time scale and then normalized to the interval [0,1].

During the simulation, the specific parameter setting for the (m, n, k) CO linear prediction part is $(5, 10, 1)$; and for the ANN prediction part, a $(10, 4, 10)$ BP neural network is implemented with pre-defined accuracy 10^{-4} . Figure 2 and 3 give the one-step-ahead prediction results of the combined approach for each trace, where the predicted traffic follows the actual traffic visually. Then the predicted time series are further quantitatively analyzed in the following two subsections.

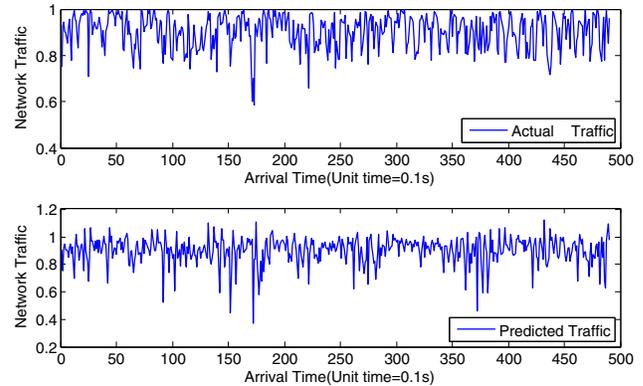


Fig. 2. One-step-ahead Prediction Results for Data I.

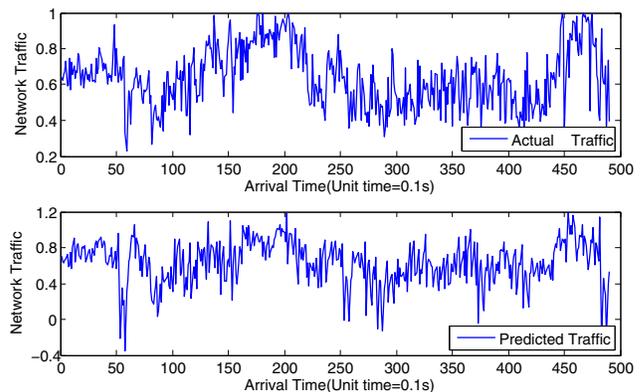


Fig. 3. One-step-ahead Prediction Results for Data II.

A. Characteristics Analysis of Predicted Traffic

The self-similarity and the burstiness of predicted traffic are evaluated through the Hurst parameter (H) and the heavy-tail index (α). The Hurst parameter is the unique measure of the self-similarity degree and $H \in (0.5, 1)$; For a self-similar process, the larger Hurst value implies the higher degree of self-similarity that the process exhibits. Besides, the characteristics of burstiness can be reflected by the heavy-tail index and $\alpha \in (1, 2)$. As α decreases, the degree of burstiness increases. Furthermore, the mean traffic level is studied due to the lack of second order moments for self-similar processes. Table I and II show the results of the above analyses, from which the following results could be got:

(1) The CO linear prediction part can effectively capture the self-similar and bursty characteristics of network traffic. From Table I and II, the Hurst parameter and the heavy-tail index of the predicted traffic in the first stage (the CO prediction stage) are increased moderately (less than 10%). Thus, the characteristics of the actual network traffic are maintained in both tests though. However, the characteristics of the final predicted traffic (under the hybrid prediction) for both tests are deeper affected than in the CO prediction. For the first trace, H and α both decrease more than 10%, implying that the self-similarity and the burstiness are seemingly changed. And for

TABLE I
COMPARISONS OF TRAFFIC CHARACTERISTICS FOR DATA I

	Original	CO	Combined
Hurst	0.733	0.807	0.542
alpha	1.873	1.880	1.541
mean	0.905	0.781	0.904

TABLE II
COMPARISONS OF TRAFFIC CHARACTERISTICS FOR DATA II

	Original	CO	Combined
Hurst	0.861	0.882	0.813
alpha	1.053	1.079	1.107
mean	0.641	0.523	0.633

the second trace, H decreases while α increases, both with a larger amount than in the CO prediction. By comparisons, this is due to the effect of the ANN prediction part. Thus the ANN adapted here shows the limitation of fitting the self-similar and bursty dynamics of network traffic in both tests.

(2) The combined prediction could maintain the mean level of the actual network traffic. The mean of the predicted traffic decreased 14% and 18% for each traffic trace respectively, which leads a seeming degradation in the prediction accuracy under the CO prediction. However, the mean of the predicted traffic varies slightly (approximate 1%) under the combined prediction. This is gained at the cost of the increased system complexity under the combined prediction. Through comparisons, the ANN prediction part accounts for the maintaining of the mean traffic level. In addition, this is mainly due to the self-learning capability of ANNs.

From the above analyses, the CO prediction part acts as a coarse prediction and could capture the dynamical behaviors of self-similar and bursty network traffic; and the ANN prediction part could maintain the mean traffic level and act as a fine prediction; however, the ANNs could affect the self-similarity and burstiness of the predicted traffic.

B. Prediction Error Analysis

To further compare the prediction accuracy, the Cumulative Error (CE) and the Absolute Cumulative Error (ACE) are used as performance metrics for prediction error analysis. The two metrics are defined as follows:

$$CE(N) = \sum_{n=1}^N [y(n+k) - x(n+k)] \quad (7)$$

$$ACE(N) = \sum_{n=1}^N |y(n+k) - x(n+k)| \quad (8)$$

where $y(n+k)$ is the predicted traffic value, $x(n+k)$ is the actual traffic value, and N is the number of predicted data. Figures 4 and 5 illustrate the CE versus the number of predicted data N for both the CO and the combined approaches. From these two figures, the CEs of both traces decrease in an approximately linear way under the CO linear prediction while remain basically at zero for those of the combined approach, independent of the increase of N , for both traces.

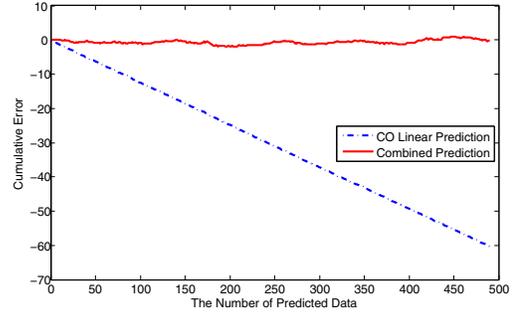


Fig. 4. Cumulative Error for Data I.

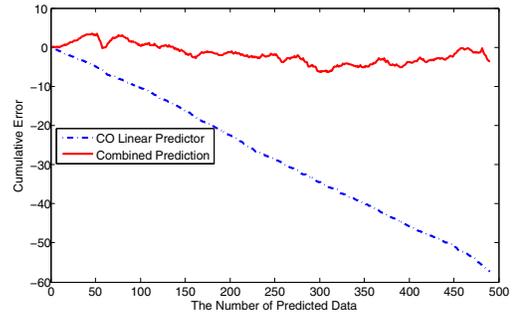


Fig. 5. Cumulative Error for Data II.

Besides, Figures 6 and 7 illustrate the ACE versus the number of predicted data N for both the CO and the combined approaches. For the first trace, the ACE of the combined approach grows slower than that of the CO approach; while the case goes opposite for the second trace. Considering that the second trace fluctuates more widely than the first trace, the capability of ANNs to fit this traffic trace is further limited. Thus, the ACE of the predicted traffic under the combined approach grows faster than under the CO prediction for the second trace.

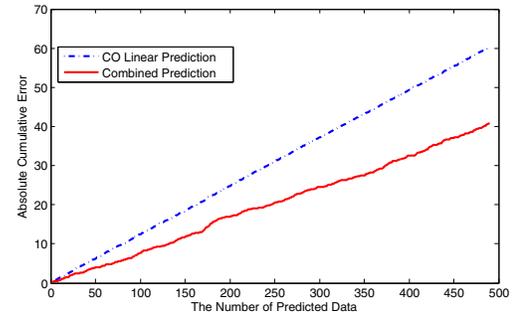


Fig. 6. Absolute Cumulative Error for Data I.

Based on the above analysis, when the real traffic fluctuates in a small range, or becomes less bursty, the combined

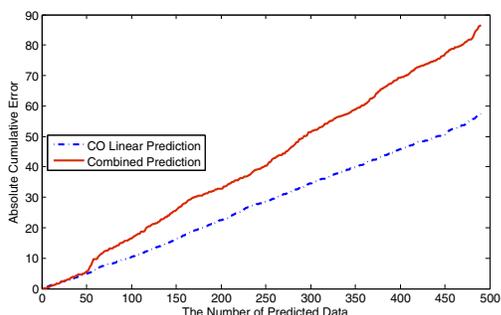


Fig. 7. Absolute Cumulative Error for Data II.

prediction could maintain the mean level of network traffic and the prediction error can be reduced seen from the CEs and the ACEs. Thus, the prediction accuracy can be effectively raised compared with the CO linear prediction. Considering the effects of both prediction parts, the CO prediction can capture the characteristics of self-similar traffic in the first stage, revealed from the changes of self-similarity and burstiness parameters; besides, the BP neural networks can finely adjust the prediction accuracy in the second stage, particularly maintaining the mean level with reduced prediction errors. This is mainly due to the self-learning capability of ANNs in the combined approach. However, the improvement of the prediction accuracy is limited when the real traffic fluctuates widely, or becomes bursty enough. In part A, the ANNs are found to affect the self-similarity and burstiness of the predicted traffic, showing limited capability in fitting real traffic. This effect is seemingly seen when the real traffic fluctuates widely enough. Moreover, it results the faster increase in the ACE of the predicted traffic for the second trace. Thus the ANNs affect the further improvement of prediction accuracy.

VI. CONCLUSION

A new hybrid prediction method combining the CO criterion and the ANNs is proposed to predict the bursty and self-similar network traffic. Through the comparison study using the real collected traffic traces, the combined prediction outperforms the CO method in the prediction accuracy, seen from the mean value level of network traffic and the prediction error. Meanwhile, the ANNs can affect the self-similarity and burstiness of the predicted network traffic. And thus the improvement in prediction accuracy becomes limited especially when the real traffic fluctuates widely enough. In the future work, we would further investigate the effect of ANNs on characteristics of the predicted network traffic and then optimize the combined prediction scheme. Besides, other forms of combination will be explored and evaluated, e.g., the tight coupling of the CO criterion in the design of the ANNs, like the combination of the CO criterion with the self-learning process.

ACKNOWLEDGMENT

The authors acknowledge the support from the National Natural Science Foundation of China (NSFC), contract/grant number: 60872007; National 863

High Technology Program of China, contract/grant number: 2009AA01Z239; The Ministry of Science and Technology (MOST), International Science and Technology Collaboration Program, contract/grant number: 0903; and the RCUK for the UK-China Science Bridges Project: R&D on (B)4G Wireless Mobile Communications. L. Shu's research in this paper was supported by Grant-in-Aid for Scientific Research (S)(21220002) of the Ministry of Education, Culture, Sports, Science and Technology, Japan. C.-X. Wang also acknowledges the support from the Scottish Funding Council for the Joint Research Institute in Signal and Image Processing with the University of Edinburgh, as part of the Edinburgh Research Partnership in Engineering and Mathematics (EPRem).

REFERENCES

- [1] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of Ethernet traffic", *IEEE/ACM Trans. Networking*, vol. 2, no. 1, 1994, pp. 1-15.
- [2] C.-X. Wang and M. Patzold, "A novel generative model for burst error characterization in Rayleigh fading channels", *Proc. 14th IEEE Int. Symp. on Personal, Indoor and Mobile Radio Communications, IEEE PIMRC 2003*, Sept. 7-10, 2003, pp. 960-964.
- [3] C.-X. Wang and M. Patzold, "A generative deterministic model for digital mobile fading channels", *IEEE Communications Letters*, vol. 8, no. 4, pp. 223-225, April 2004.
- [4] C.-X. Wang and W. Xu, "A new class of generative models for burst error characterization in digital wireless channels", *IEEE Trans. Communications*, vol. 55, no. 3, pp. 453-462, March 2007.
- [5] M. Lopez-Guerrero, L. Orozco-Barbosa, and D. Makrakis, "Probabilistic envelope processes for alpha-stable self-similar traffic models and their application to resource provisioning", *Performance Evaluation*, vol. 61, no. 2-3, pp. 257-279, Jul. 2005.
- [6] J.R. Gallardo, D. Makrakis, and L. Orozco-Barbosa, "Prediction of alpha-stable long-range dependent stochastic processes", in *IEEE/IEE Int. Conf. Telecommunications*, vol.2, pp. 222-226, Cheju, Korea, Jun. 1999.
- [7] Y. Wen, G. Zhu, C. Xie, "Prediction for wireless traffic with long range burstiness", in *Circuits and Systems for Communications, 2008. ICCSC 2008. 4th IEEE Int. Conf.*, pp.461-465, May 2008.
- [8] X. Ge, S. Yu, W. Yoon, and Y. Kim, "A new prediction method of alpha-stable processes for self-similar traffic", in *Global Telecommunications Conference, 2004. GLOBECOM '04. IEEE*, vol.2, pp. 675- 679, 2004.
- [9] M. Lopez-Guerrero, J.R. Gallardo, D. Makrakis, and L. Orozco-Barbosa, "Sensitivity of a network traffic prediction algorithm", *Communications, Computers and signal Processing, 2001. PACRIM. 2001 IEEE Pacific Rim Conference on*, vol. 2, pp. 615-618, 2001.
- [10] M. Lopez-Guerrero, J.R. Gallardo, D. Makrakis, and L. Orozco-Barbosa, "Optimizing linear prediction of network traffic using modeling based on fractional stable noise", *Info-tech and Info-net, 2001. Proceedings. ICII 2001 - Beijing. 2001 Int. Conf.*, vol. 2, pp. 587-592, 2001.
- [11] M. Lopez-Guerrero, J.R. Gallardo, L. Orozco-Barbosa, and D. Makrakis, "On the dynamic allocation of resources using linear prediction of aggregate network traffic", *Computer Communications*, vol. 26, no. 12, pp. 1341-1352, July 2003.
- [12] H. Zhao, "Multiscale analysis and prediction of network traffic", *Performance Computing and Communications Conference (IPCCC), 2009 IEEE 28th International*, pp. 388-393, Dec. 2009.
- [13] L. Zhu, L. Qin, K. Xue, and X. Zhang, "A novel BP neural network model for traffic prediction of next generation network", in *Natural Computation, 2009. ICNC '09. Fifth Int. Conf.*, vol. 1, pp.32-38, Aug. 2009.
- [14] Y. Liu, and J. Cao, "Network traffic prediction based on grey neural network integrated model", in *Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08*, pp.1-4, Oct. 12-14, 2008.
- [15] The CONAN Project: <http://moment.cs.ucsb.edu/conan/>
- [16] X. Ge, Y. Yang, C.-X. Wang, Y.-Z. Liu, C. Liu, and L. Xiang, "Characteristics analysis and modeling of frame traffic in 802.11 wireless networks", *Wireless Communications and Mobile Computing, John Wiley & Sons*, vol. 10, no. 4, pp. 584-592, Apr. 2010.