International Journal of Multimedia and Ubiquitous Engineering
Vol. 12, No. 5 (2017), pp. 29-42
http://dx.doi.org/10.14257/ijmue.2017.12.5.03

# Predicting Bursty Network Traffic with Self-similarity Characteristic over Echo State Covariation Orthogonality Network

Xiaochuan Sun[1*], Yingqi Li[1] and Minghui Zhang[1]

*[1]North China University of Science and Technology, Tangshan 063009, China*
*[*]xiaochuansun@yeah.net*

***Abstract***

*Network traffic depicts the network characteristics and users behaviors. Accurate network traffic prediction is essential for dynamic network management. In this paper, the echo state covariation orthogonality network (ESCON) is proposed in a linear unbiased estimation framework based on echo state mechanisms for network traffic prediction. The ESCON inherits the basic idea of ESN learning in an unbiased estimation framework, but replaces the commonly used least square method with a covariation orthogonality one, which can reflect the tendency of network traffic more accurately, to solve the optimal output weights. We perform a comprehensive performance evaluation, considering publicly available nonstationary H.264 video traces. In all traces, we show that the ESCON can more effectively capture the characteristics of self-similarity and bursty, and yield superior prediction accuracy than the considered prediction schemes.*

***Keywords****: echo state network; covariation orthogonality; self-similarity; bursty*

## 1. Introduction

Network traffic dramatically affects the network activities, such as network management, traffic engineering, network protocol analysis, network design, routing optimization. In the recent years, network traffic has experienced significant changes with the development of video conference, high-definition TV, video browsing applications and virtual reality application. Some of emerging applications require strict quality of service (QoS) guarantee on packet loss, delay and delay jitter. In order to satisfy the QoS requirements, accurate traffic modeling and prediction [1][2] are required for adequate call admission control, usage parameter control, traffic management, resource allocation and congestion control [3][4]. Recently, with the rapid development of the communication and network technologies, the traffic characteristics have changed drastically. Network traffic [5] over high-speed Internet typically can possess the characteristics self-similarity and long range dependence, and simultaneously exhibit high bursty over different time scales. It is a challenging problem to measure the network traffic directly in real networks [6]. Network traffic modeling and prediction have become a research hotspot and draw widespread attention.

To the best of our knowledge, the neural network (NN) is one of the most usually used nonlinear methods to predict traffic data. Park [7] proposed a structurally optimized bilinear recurrent neural network that had been successfully applied to the network traffic prediction. Cortez *et al*. [8] developed a fully connected multilayer perceptron (MLP) to predict the amount of traffic in TCP/IP based networks. Katris and Daskalaki [9] evaluated the fractal autoregressive integrated moving average (FARIMA) model, MLP and radial basis function (RBF) neural network, and further proposed the two alternative approaches that could combine the merits of both FARIMA and neural network models. Oravec *et al*. [10] proposed video traffic prediction methods based on the MLP, RBF and backpropagation through time (BPTT) neural networks to predict the

ISSN: 1975-0080 IJMUE

variable bit rate video (VBR) traffic for the efficient bandwidth allocation. Ding *et al*. [11] proposed a wavelet transformation based integrated prediction model to deal with the dynamic characteristics of network traffic. Zhang *et al*. [12] proposed a fuzzy wavelet neural network (WNN) based on the quantum-behaved particle swarm optimization for network traffic to extract the mapping relations between the input and output data. Although the above traditional neural networks (e.g. MLP, RBF, BPTT and WNN) have the nonlinear approximation capability, they also exhibit many weaknesses, characterized by slow convergence, over-fitting phenomenon and suboptimal solution.

As a powerful neural network, echo state network (ESN) [13][14][15] has the characteristics of a remarkable dynamic reservoir and simple linear readout. The reservoir consists of massive randomly and sparsely connected neurons. The unique trainable part is the output weights that which can be solved through simple linear regression. It is proved that the ESN outperform the traditional neural networks. The ESN has been applied successfully in speech recognition, reinforcement learning, robot control and time-series prediction. However, we have verified that it is rather difficult for the ESN based on simple regression method (e.g. pseudo-inverse method or ridge regression method) to accurately predict network traffic. Fortunately, Xiang *et al*. [16] introduced the covariation orthogonality (CO) criterion into network traffic prediction, which provided a superior linear unbiased estimation method. The paper focuses on the combination of the ESN and CO criterion to describe the prominent characteristics of network traffic as accurately as possible.

In this paper, we propose an echo state covariation orthogonality network (ESCON) for network traffic prediction, combining the merits of the CO criterion and echo state mechanisms. The basic idea of ESCON is to train ESN in a linear unbiased estimation framework, while replacing the least square method with the CO criterion, which is less sensitive to outliers and able to effectively capture the self-similarity and bursty of network traffic. Theoretically, we derive the output weights formula available for ESCON. Besides, we also consider the multistep-ahead prediction for network traffic, since network bandwidth allocation/reallocation can never be achieved in such small time scale. The efficacy of the proposed approach is evaluated on nonstationary H.264 video traces, using a well-known video trace dataset. Its performance is compared to Gaussian process based ESN [15], a combined method (back propagation neural network & CO criterion) [16], and multiresolution-learning-based NN predictor [17].

The rest of the paper is organized as follows. Section 2 introduces the linear unbiased estimation based on CO criterion. Section 3 describes the ESCON in detail, and derives the output weight matrix. Section 4 shows the experimental results and performance analysis for real-world network traffic. Finally, section 5 gives the short conclusion about our work.

## 2. Covariation Orthogonality Criterion

Consider that the covariance is used to effectively conduct the Gaussian random process. However, in the $\alpha$-stable process [18], whenever the characteristic parameter $\alpha < 2$, the population variance is proved to be infinite, resulting in ineffectiveness of the covariance function. To gain a better understanding of $\alpha$-stable process, the covariation has been proposed to replace the covariance in the case of $\alpha < 2$ [18]. The covariation can be defined as follow:

**Definition 1(Covariation Orthogonality)[16]:** Assume that $X$ and $Y$ are random vectors of the $\alpha$-stable process with $\alpha > 1$, and for the polar coordinates space $\boldsymbol{R}^2$, $(u, v)$ is the representation of the process $(X, Y)$ with spectral measure $\boldsymbol{\Gamma}$ on the unit circle $S$, the covariation of $X$ on $Y$ is given by:

$$[X,Y]_\alpha = \int_S uv^{\langle \alpha-1 \rangle} d\Gamma(u,v) \tag{1}$$

where <·> denotes the sign power, given by

$$\alpha^{\langle p \rangle} = |\alpha|^p \, sign\alpha = \begin{cases} \alpha^p & \alpha \geq 0 \\ -|\alpha|^p & \alpha < 0 \end{cases} \tag{2}$$

When the $\alpha$-stable random vectors $X$ and $Y$ satisfy the following formula

$$[X,Y]_\alpha = 0 \tag{3}$$

the vector $X$ is regarded as the covariation orthogonality related to the vector $Y$.

Definition 2(Unbiased Estimation): Let ($X_1$, $X_2$, …, $X_n$; $Y_n$) be joint alpha-stable distribution, and $\hat{Y}_n$ is the estimated value of $Y_n$ , if satisfy:

$$E[\varepsilon_n | X_1, X_2, ..., X_n] = 0 \tag{4}$$

where $\varepsilon_n = Y_n - \hat{Y}_n$, $\hat{Y}_n$ is said to be unbiased estimator to the $Y_n$.

**Lemma 1**[16]**:** Let the vector ($X_1$, $X_2$, …, $X_n$; $Y_n$) be the joint $\alpha$-stable distribution with the characteristic parameter $\alpha\in(1, 2)$, and for the spectral measure $\boldsymbol{\Gamma}$, the ($v$, $\boldsymbol{u}$) is the spectral representation of ($Y_n$, $\boldsymbol{X}$), yield:

A) For the set $\{a_i\}_{i=1}^m$, $E[Y_n|\mathbf{X}] = \sum_{i=1}^{m} a_i X_{n+1-i}$ , if and only if $\forall \mathbf{t} \in \Re^m$ ,

$$\int_{S_{m+1}} (v - \mathbf{a}'\mathbf{u})(\mathbf{t}'\mathbf{u})^{\langle \alpha-1 \rangle} d\Gamma(v, \mathbf{u}) = 0 \tag{5}$$

B) The following formulas are equivalent.

$$E[Y_n|\mathbf{X}] = \mathbf{a}'\mathbf{X} \tag{6}$$

$$E[Y_n - \mathbf{a}'\mathbf{X}|\mathbf{X}] = 0 \tag{7}$$

$$[Y_n - \mathbf{a}'\mathbf{X}, \mathbf{X}]_\alpha = 0 \tag{8}$$

According to the **Lemma 1**, for the joint alpha-stable ($X_1$, $X_2$, …, $X_n$; $Y_n$), observe that $Y_n$ is a linear unbiased estimation based on the CO criterion for the $\mathbf{a}'\mathbf{X}$ .

## 3. ESCON

### 3.1. ESN in a Nutshell

The ESN [19] is a special form of recurrent neural network, which is trained through the supervised learning. The general structure of an ESN is shown in Figure 1. Observe that the network consists of $K$ input units $u(t)=(u_1(t), u_2(t), \ldots, u_K(t))$, $N$ internal units $x(t)=(x_1(t), x_2(t),\ldots, x_N(t))$, and $L$ output units $y(t)=(y_1(t), y_2(t),\ldots, y_L(t))$. The real-valued connection weights are obtained in a $N\times K$ weight matrix $W^{in}$ for the input weights, in a $N\times N$ weight matrix $W$ for the internal connections, in a $L\times N$ weight matrix $W^{out}$ for the connections from internal units to the output units. Each neuron or unit of the ESN has an activation state at a given time step $t$. When the input stream $u(t)$ is fed into the reservoir units, the activation states of the reservoir units are generally updated using the following equation:

$$x(t+1) = f(W^{in}u(t+1) + Wx(t)) \tag{9}$$

where $f$ is the transfer function of the reservoir units (*tanh*, other sigmoidal function or identity function). The generic output is calculated according to the system internal state by

$$y(t+1) = f^{out}(W^{out}x(t+1)) \tag{10}$$

where the output function $f^{out}$ is usually linear.

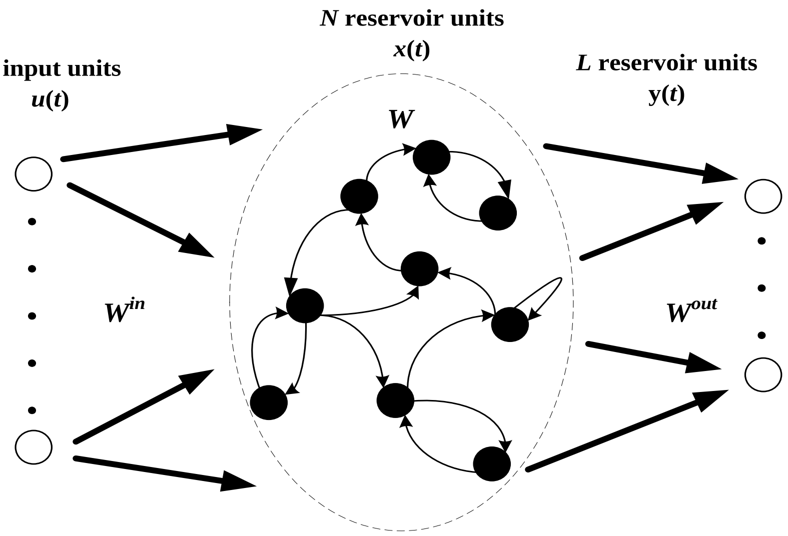



**Figure 1. The General Structure of an ESN**

In the standard ESN implementations, $W^{in}$ and $W$ remain unchanged after initialization, and the spectral radius of the reservoir weight matrix $W$ is set less than 1 to obtain the echo state property.

### 3.2. ESCON Formulation

In this section we put forward the theoretical consideration on how the CO criterion is introduce for the ESN training. We formalize the interaction between the reservoir dynamics and training error, and further solve the optimal output weights. To the best of our knowledge, the ESN using pseudo-inverse method or ridge regression method can cause the significant degradation of nonlinear approximation capability when facing the network traffic with self-similarity and bursty characteristics.

The following quantitative analysis can give deeper insights into the training method of ESN. Assume that $x(t)=(x_1(t), x_2(t),\ldots, x_N(t))$ denotes the reservoir states, $y(t)=(y_1(t), y_2(t),\ldots, y_L(t))$ denotes the desired output and $\hat{y}(t)$ is the predicted output at the time step $t$. According to Eq. (10), the predicted output is given by

$$\hat{y}(t) = W^{out} x(t) \tag{11}$$

We can observe from Eq. (11) that only after the output weight matrix $W^{out}$ is solved, can the predicted output $\hat{y}(t)$ be obtained. Let the actual output $y(t)$ and reservoir states $x(t)$ be joint $\alpha$-stable process. According to the Definition 2, there exists a linear unbiased estimation based on the CO criterion, yielding

$$E[y(t) - \hat{y}(t) | x(t)] = 0 \tag{12}$$

Moreover, we have (see **Lemma 1**)

$$[y(t)\text{-}W^{out}x(t), x(t)]_a = 0 \tag{13}$$

where the prediction error $\varepsilon=y(t)\text{-}W^{out}x(t)$ is covariation orthogonality to the reservoir state $x(t)$. Based on the decomposability of covariation [18], we can further convert the Eq. (13) to

$$[y(t), x(t)]_a \text{-} [W^{out}x(t), x(t)]_a = 0 \tag{14}$$

Since the first vector of covariation equation is linear, we have

$$W^{out}[x(t),x(t)]_a = [y(t),x(t)]_a \tag{15}$$

where $[x(t),\ x(t)]_\alpha$ denotes the self-covariation, and $[y(t),\ x(t)]_\alpha$ denotes the cross-covariation. The Eq. (15) can be written as the following covariation matrix form

$$W^{out}\cdot\begin{bmatrix} [x_N(t),x_N(t)]_\alpha & [x_{N-1}(t),x_N(t)]_\alpha & \cdots & [x_1(t),x_N(t)]_\alpha \\ [x_N(t),x_{N-1}(t)]_\alpha & [x_{N-1}(t),x_{N-1}(t)]_\alpha & \cdots & [x_1(t),x_{N-1}(t)]_\alpha \\ \vdots & \vdots & \vdots & \vdots \\ [x_N(t),x_1(t)]_\alpha & [x_{N-1}(t),x_2(t)]_\alpha & \cdots & [x_N(t),x_N(t)]_\alpha \end{bmatrix} = \begin{bmatrix} [y_L(t),x_N(t)]_\alpha & [y_L(t),x_{N-1}(t)]_\alpha & \cdots & [y_L(t),x_1(t)]_\alpha \\ [y_{L-1}(t),x_N(t)]_\alpha & [y_{L-1}(t),x_{N-1}(t)]_\alpha & \cdots & [y_{L-1}(t),x_1(t)]_\alpha \\ \vdots & \vdots & \vdots & \vdots \\ [y_1(t),x_N(t)]_\alpha & [y_1(t),x_{N-1}(t)]_\alpha & \cdots & [y_1(t),x_1(t)]_\alpha \end{bmatrix} \tag{16}$$

Under the assumption of $(X,\ Y)$ obeying joint alpha-stable distribution with $\alpha>1$, the covariation has the following form [16]:

$$[X,Y]_\alpha = \frac{\sigma_Y^{\alpha-p}}{E|X_0|^p} EXY^{<p-1>}, 1<p<\alpha \tag{17}$$

where $\sigma_Y$ is covariation norm of vector $Y$. Similarly, we have

$$[X,X]_\alpha = \frac{\sigma_X^{\alpha-p}}{E|X_0|^p} EXX^{<p-1>} \tag{18}$$

$$[X,X]_\alpha = \frac{\sigma_X^{\alpha-p}}{E|X_0|^p} EYX^{<p-1>} \tag{19}$$

The above two formulas are substituted into Eq. (16), and after some manipulation, the output weight matrix of ESCON is given by

$$W^{out} = \begin{bmatrix} E[y_L(t)x_N^{<p-1>}(t)] & E[y_L(t)x_{N-1}^{<p-1>}(t)] & \cdots & E[y_L(t)x_1^{<p-1>}(t)] \\ E[y_{L-1}(t)x_N^{<p-1>}(t)] & E[y_{L-1}(t)x_{N-1}^{<p-1>}(t)] & \cdots & E[y_{L-1}(t)x_1^{<p-1>}(t)] \\ \vdots & \vdots & \vdots & \vdots \\ E[y_1(t)x_N^{<p-1>}(t)] & E[y_1(t)x_{N-1}^{<p-1>}(t)] & \cdots & E[y_1(t)x_1^{<p-1>}(t)] \end{bmatrix} \cdot \begin{bmatrix} E[x_N(t)x_N^{<p-1>}(t)] & E[x_{N-1}(t)x_N^{<p-1>}(t)] & \cdots & E[x_1(t)x_N^{<p-1>}(t)] \\ E[x_N(t)x_{N-1}^{<p-1>}(t)] & E[x_{N-1}(t)x_{N-1}^{<p-1>}(t)] & \cdots & E[x_1(t)x_{N-1}^{<p-1>}(t)] \\ \vdots & \vdots & \vdots & \vdots \\ E[x_N(t)x_1^{<p-1>}(t)] & E[x_{N-1}(t)x_1^{<p-1>}(t)] & \cdots & E[x_1(t)x_1^{<p-1>}(t)] \end{bmatrix}^{-1} \tag{20}$$

where for $\forall i = 0,1,....,N-1,\ j = 0,1,...,N-1$ , the population mean of $x_{t-i}(t)x_{t-j}^{<p-1>}(t)$, the $E[x_{t-i}(t)x_{t-j}^{<p-1>}(t)]\ \ t = 1,2,\cdots,N$ can be evaluated as follow

$$E[x_{t-i}(t)x_{t-j}^{<p-1>}(t)] = \frac{1}{N+min\{i,\ j\}-max\{i,\ j\}} \sum_{t=max\{i,\ j\}+1}^{N+min\{i,\ j\}} x_{t-i}(t)x_{t-j}^{<p-1>}(t) \tag{21}$$

By analogy, for $\forall i = 0,1,\cdots,L-1,\ j = 0,1,\cdots,N-1$, the $E[y_{s-i}(t)x_{s-j+N-L}^{<p-1>}(t)]$ is given by

$$E[y_{s-i}(t)x_{s-j}^{<p-1>}(t)] = \frac{1}{L+\min\{i,j\}-\max\{i,j\}} \sum_{s=\max\{i,j\}+1}^{L+\min\{i,j\}} y_{s-i}(t)x_{s-j+N-L-\min\{i,j\}}^{<p-1>}(t) \tag{22}$$

Finally, employing Eq. (21) and (22), the optimal output weight matrix $W^{out}$ of the ESCON can be calculated by solving Eq. (20), which implies that the ESCON training is finished by the linear unbiased estimation based on CO criterion.

**Table 1. H.264 Video Traces**

| Video trace | Coding standard | Type |
|---|---|---|
| IMAX Station Space | MVC | Movie |
| Tokyo Olympics | AVC | Sports |
| News | SVC | TV News |

**Table 2. Experimental Configuration for Three H.264 Video Traffic**

| Parameters | IMAX station space | Tokyo olympics | News |
|---|---|---|---|
| Reservoir size | 80 | 80 | 50 |
| Spectral radius | 0.8 | 0.95 | 0.8 |
| Reservoir connectivity | 0.1 | 0.1 | 0.1 |
| Output feedback | No | No | No |
| Washout time | 500 | 500 | 1500 |

## 4. Experiments

In this section, we provide a thorough experimental evaluation of the ESCON, considering three video traffic traces with H.264 coding [20]. More detail traffic characteristics and coding parameters of the H.264 video traces can be found in [21]. In our experiment, the used live real-world H.264 video traffic traces include the 3D movie, sport and news, covering a wide range of popular video samples, as listed in Table 1. To verify the potential of ESCON, we analyze the characteristic changes (i.e. self-similarity and bursty) of H.264 video traffic before and after prediction, and evaluate the prediction performance of proposed ESCON. Compared with echo state Gaussian process [15] (ESGP), a combined method (BP neural network & covariation orthogonality prediction, CO-BPNN) [16], and multiresolution-learning-based NN predictor (MLNN, in my case, the three layer NN is 24-5-1, and Haar wavelet is used for multiresolution analysis.) [17]. The experimental settings for the different H.264 video traces are shown in Table 2. In order to simplify the calculation, the parameter $p$ has the same value as the parameter $\alpha$ .

To quantitatively assess the prediction performance, the normalized mean squared error (*NMSE*) and the prediction gain $R_p$ are used, which are respectively defined as

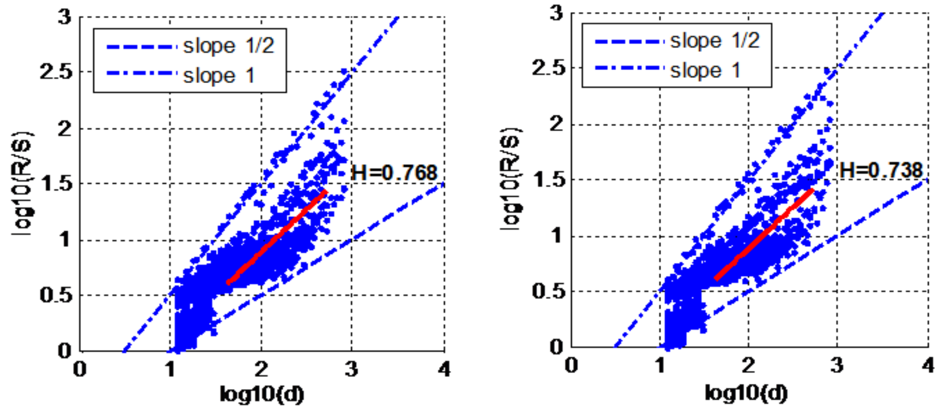
$$NMSE = \frac{\sum_{i=1}^{l_{test}} (\hat{y}(t) - y(t))}{l_{test} \cdot \sigma^2} \tag{23}$$

where $\hat{y}(t)$ and $y(t)$ are the predicted output and actual output respectively during the testing phase, and $\sigma^2$ is the variance of the actual output over the prediction duration $l_{test}$.

$$R_p = 10\log_{10}(\frac{\sigma_u^2}{\sigma_e^2})\mathrm{dB} \tag{24}$$

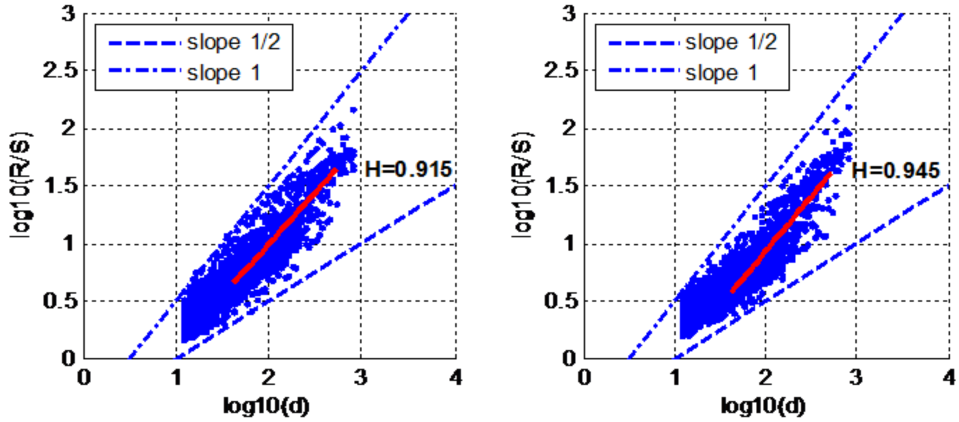
where $\sigma_u^2$ denotes the estimated variance of the input, and $\sigma_e^2$ denotes the estimated variance of the prediction error.



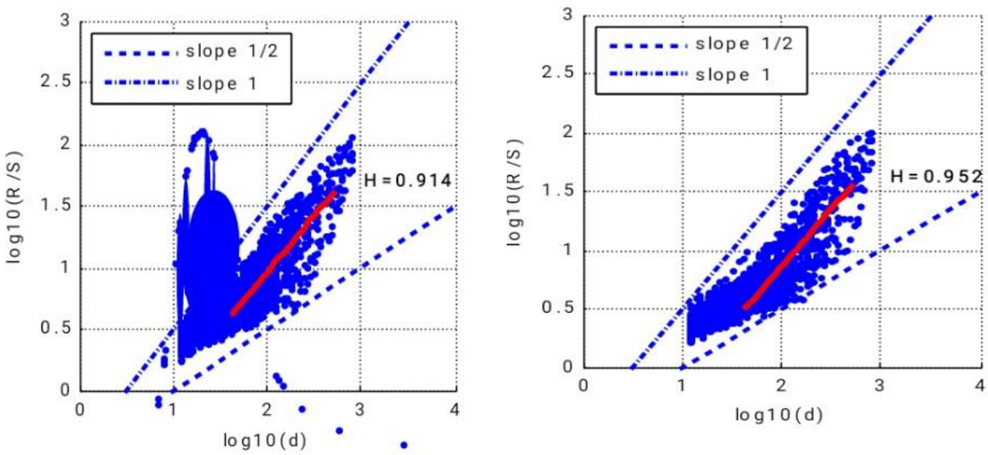

(a) (b)

**Figure 2. R/S Plots of 200-Frame-Ahead Prediction over the ESCON for IMAX Station Video. (a) Actual Video Traffic, (b) Predicted Video Traffic**




(a) (b)

**Figure 3. R/S Plots of 200-Frame-Ahead Prediction over the ESCON for Tokyo Olympics Video. (a) Actual Video Traffic, (b) Predicted Video Traffic**




(a) (b)

**Figure 4. R/S Plots of 200-Frame-Ahead Prediction over the ESCON for News Video. (a) Actual Video Traffic, (b) Predicted Video Traffic.**

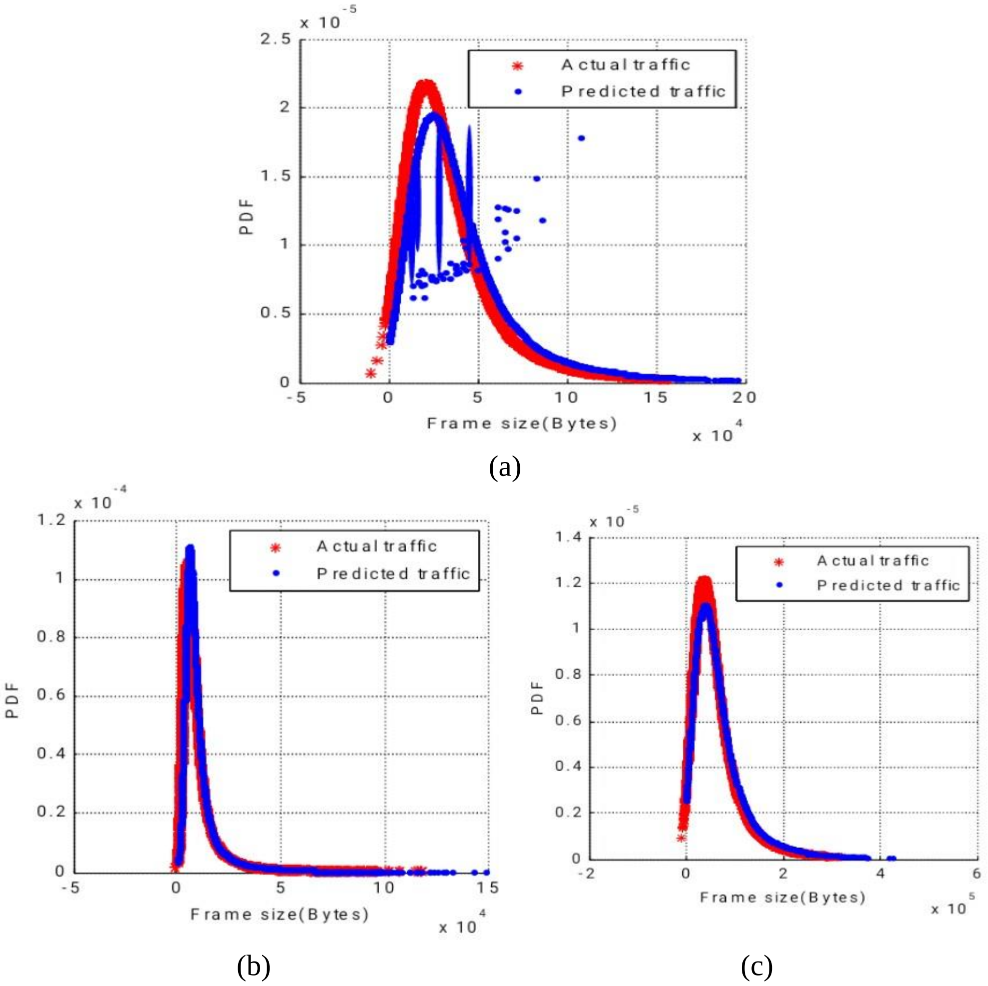
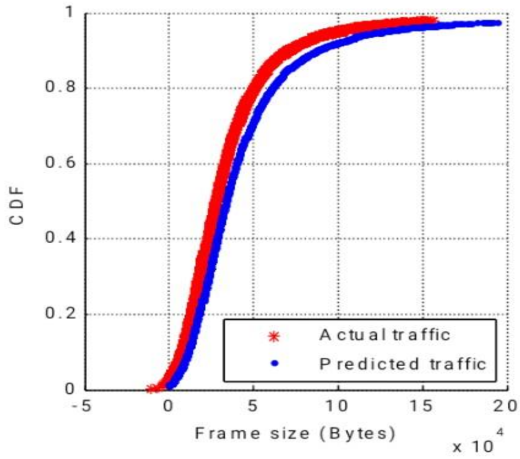



**Figure 5. PDF Fitting Plots of Actual and Predicted Traffic for Three H.264 VideoTraces over the 200-Frame-Ahead Prediction of ESCON. (a) IMAX StationSpce. (b) Tokyo Olympics. (c) News**

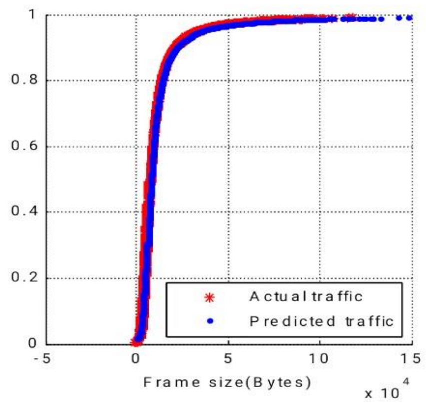
### 4.1. Characteristic Comparison of H.264 Video traces

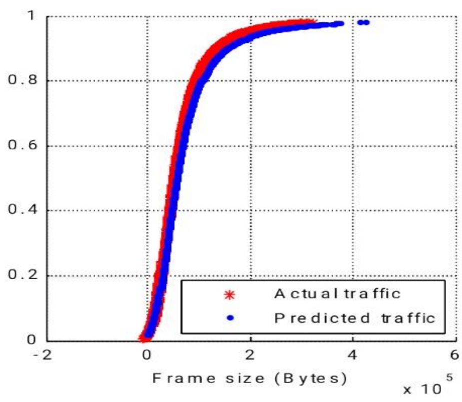
Network traffic has strong self-similarity, which is measured by the Hurst parameter $H$. $H \in (0.5, 1)$ [23] illustrates the existence ofself-siilarity, and the larger $H$, the stronger self-similarity degree. Moreover, the self-similarity can lead to heavy-tail distribution. Based on the $R/S$ method [23], Figure. 2-4 respectively give the comparisons self-similar degrees of actual and predicted traffic for the three H.264 video traces over the ESCON, where the slope of the red line is the value of Hurst parameter. It can see directly that the three actual H.264 video traces exhibit strong self-similarity, whose values of Hurst parameters are 0.768, 0.915 and 0.914, respectively. We also observe that the values of Hurst parameters, obtained before and after the ESCON is applied, are almost the same. This indicates that the proposed ESCON can effectively model the selfsimilarity of network traffic.

Moreover, network traffic also exhibits high bursty, characterized by the heavy-tail index $\alpha$. $\alpha \in (1, 2)$ [10] dominates the bursty degree of network traffic, and the smaller $\alpha$ reflects the higher bursty. Figure 5 shows the good fitting degree of the probability density functions (PDF) for the actual and predicted video traffic over the ESCON. Observe that the predicted video traffic has a more obvious heavy-tail than the actual one over the frame size for each video trace, which implies stronger bursty. In addition, the cumulative distribution function (CDF) is used to describe the distribution characteristic of the frame size. As shown in Figure 6, good fitting results of the CDFs are also obtained for the actual and predicted H.264 video traffic. The results imply that the proposed ESCON can effectively describe the bursty characteristic of the network traffic.




(a)



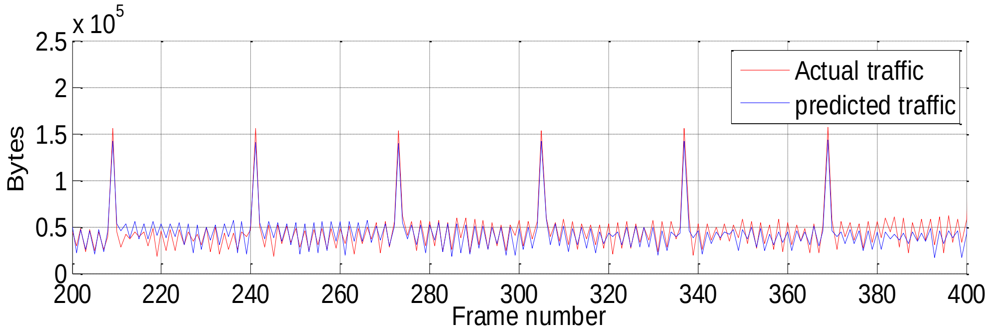

(b) (c)

**Figure 6. CDF Fitting Plots of Actual and Predicted Traffic for Three H.264 Video Traces over the 200-Frame-Ahead Prediction of ESCON. (a) IMAX Station Space, (b) Tokyo Olympics, (c) News**

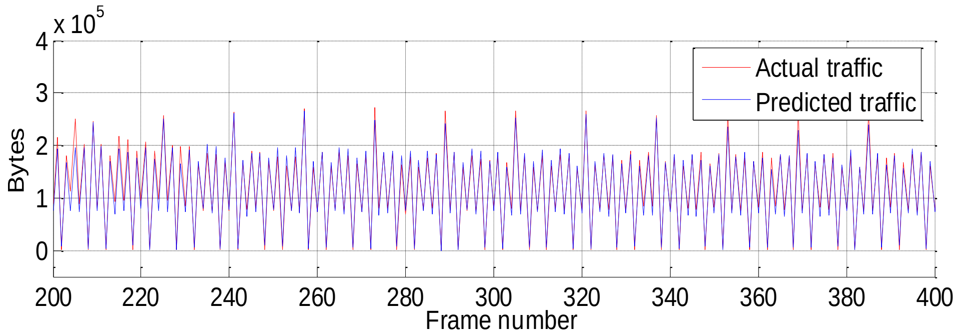
Table 3 quantitatively depicts characteristic comparisons of the actual and predicted video traffic for the different evaluated methods. As we observe, the predicted video traffic, obtained by the ESCON, shows the biggest $H$ and smallest $\alpha$ among the evaluated methods, which means that the ESCON slightly affects the self-similarity and bursty of H.264 video traffic. It is because both processes, which the video traffic is identically mapped to the high-dimensional activation states for the ESCON learning, and the CO unbiased estimation can enhance the nonlinear approximation capability, can guarantee the accurate characteristic transfer of video traffic. By contrast, ESGP, CO-BPNN and ML-NN enormously affect the dynamic characteristics of video traffic, resulting in the significant decrease of the self-similarity and bursty. Hence, these observations collectively demonstrate that the ESCON can effectively capture and maintain the dynamic behaviors of video traffic with self-similarity and bursty.

**Table 3. Characteristic Comparison of Three H.264 Video Traces Over the 200-Frame-Ahead Prediction of Evaluated Methods**

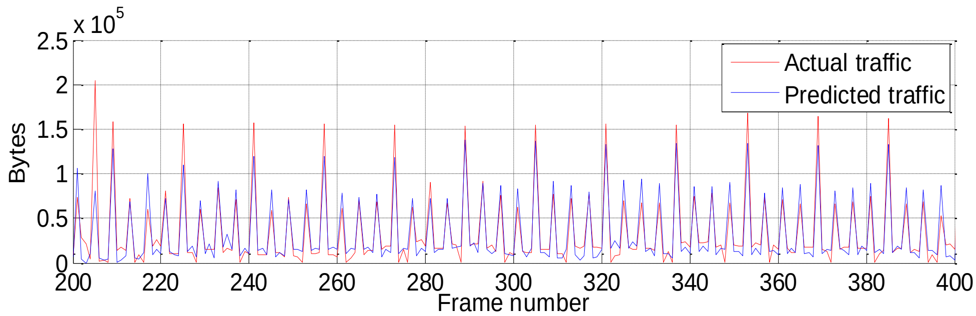
| Video trace | Parameter | Actual video traffic | ESCON | CO-BPNN | ESGP | ML-NN |
|---|---|---|---|---|---|---|
| IMAX Station Space | Hurst | 0.738 | 0.768 | 0.548 | 0.559 | 0.548 |
| | $\alpha$ | 1.376 | 1.287 | 1.603 | 1.556 | 1.536 |
| Tokyo Olympics | Hurst | 0.915 | 0.945 | 0.837 | 0.869 | 0.809 |
| | $\alpha$ | 1.098 | 1.040 | 1.261 | 1.283 | 1.193 |
| News | Hurst | 0.914 | 0.952 | 0.816 | 0.846 | 0.867 |
| | $\alpha$ | 1.316 | 1.260 | 1.451 | 1.408 | 1.384 |




(a)




(b)




(c)

**Figure 7. 200-Frame-Ahead Predictions of the ESCON for Three H.264 Video Traces. (a) IMAX Station Space (b) Tokyo Olympics (c) News**

### 4.2. Prediction Performance

In Figure 7, we give the comparative curves of the actual and predicted outputs for the ESCON, CO-BPNN, ESGP and ML-NN over the considered test region, respectively. It is clear that the proposed ESCON can achieve the good peaks-fitting for each video traces. Figure 8 shows the *NMSE*s of multi-frame-ahead prediction for three video traces. It can be seen that the proposed ESCON obviously outperforms the CO-BPNN, ESGP and ML-NN as the multi-frame step increases. It should not be surprising since the dynamic reservoir of ESCON can extract the mapping relations between the input and output data more effectively, and the unbiased estimation based on the CO criterion can ensure the approximation capacity. However, the ESGP exhibits bad prediction performance, because the inherent Gaussian process can't effectively describe the video traffic with strong self-similarity and bursty. As for the CO-BPNN and ML-NN, the backpropagation learning algorithm, owning the instability of learning and memory, is easy to trap in local optimum, resulting in the unsatisfactory prediction performance. In addition, as a standard criterion to assess prediction performance, the prediction gain $R_p$ is used to measure how well the evaluated methods perform. Table 4 shows the prediction gain variation of evaluated methods over multi-frame-ahead prediction. Observe that the ESCON obtains higher prediction gain compared with the CO-BPNN, ESGP and ML-NN, even when 500-frame-ahead prediction is performed. Hence, the ESCON is endowed with superior prediction performance, more suitable for network traffic prediction.

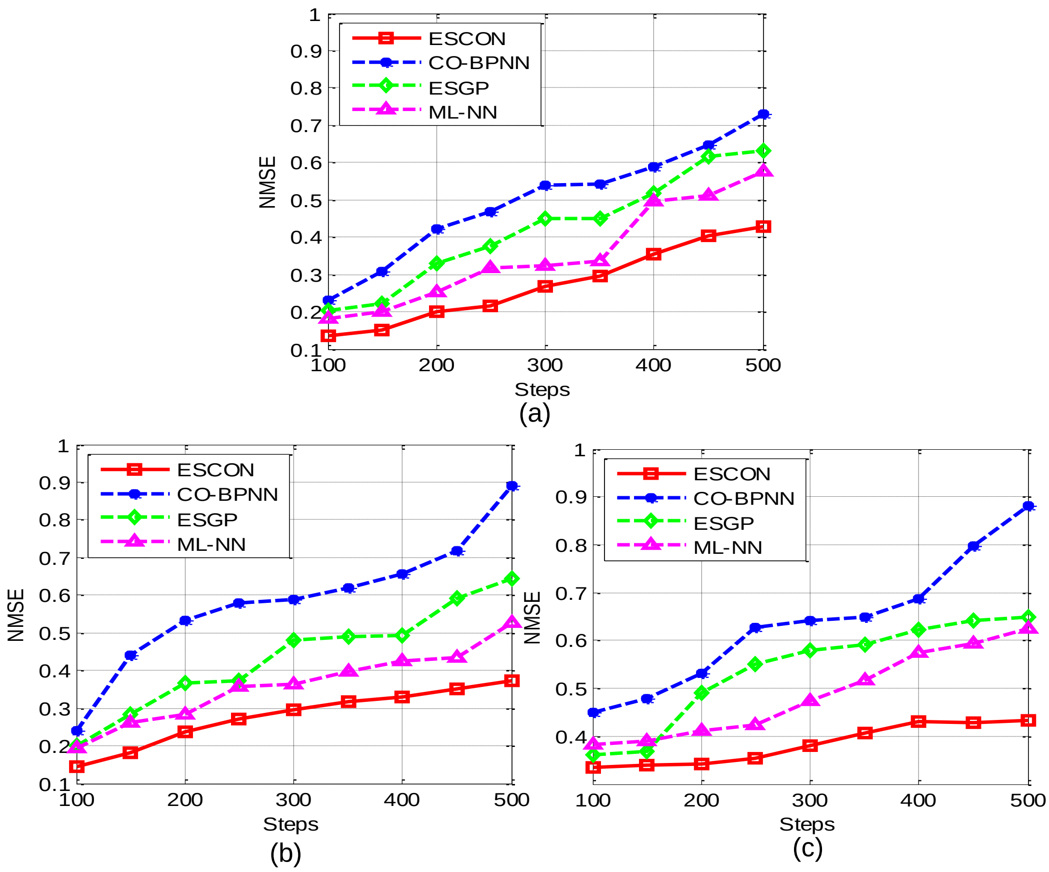



**Figure 8. Multiframe-Ahead Prediction *NMSE* of Evaluated Methods for Three H.264 Video Traces: (a) IMAX Station Space, (b) Tokyo Olympics, (c) News**

## 5. Conclusion

In this paper, we propose a new methodology for network traffic prediction, named the ESCON. The ESCON can be treated under the linear unbiased estimation framework, giving rise to an efficient model estimation algorithm based on the CO criterion. Through solving the covariation orthogonality equation of prediction error relative to the reservoir state, the so-obtained methodology can find the optimal output weights for the ESCON learning. The performance of our method is evaluated by means of simulations with three H.264 video traces. Our results indicate that the proposed ESCON exhibits significantly superior multi-frame-ahead prediction performance, and more effectively capture the self-similarity and bursty of network traffic, compared with the CO-BPNN, ESGP and ML-NN. In our future work, the open issues that we will endeavor to address, includes improving the reservoir structure and optimizing the multiple parameters of ESCON.

**Table 4. The Gain $R_p$ of Evaluated Methods for Three H.264 Video Traces**

| Video trace | Step | ESCON | CO-BPNN | ESGP | ML-NN |
|---|---|---|---|---|---|
| IMAX Station Space | 100 | 8.271 | 6.302 | 7.018 | 7.418 |
| | 200 | 6.985 | 3.708 | 4.925 | 6.157 |
| | 300 | 5.734 | 2.715 | 3.476 | 4.974 |
| | 400 | 4.594 | 2.316 | 2.905 | 3.198 |
| | 500 | 3.788 | 1.536 | 2.042 | 2.568 |
| Tokyo Olympics | 100 | 8.164 | 6.229 | 7.096 | 7.367 |
| | 200 | 6.315 | 2.758 | 4.325 | 5.439 |
| | 300 | 5.268 | 2.316 | 3.257 | 4.457 |
| | 400 | 4.802 | 1.897 | 3.051 | 3.724 |
| | 500 | 4.251 | 0.639 | 1.992 | 2.806 |
| News | 100 | 4.812 | 3.475 | 4.483 | 4.201 |
| | 200 | 4.781 | 2.774 | 3.015 | 3.882 |
| | 300 | 4.409 | 2.016 | 2.409 | 3.263 |
| | 400 | 3.912 | 1.801 | 2.157 | 2.425 |
| | 500 | 3.971 | 0.739 | 1.921 | 2.079 |

## Acknowledgments

This work is supported in part by the National Key Technology Support Program (No. 2013BAH72B01).

## References

[1] Y. Chen, B. Yang and Q. Meng, "Small-time scale network traffic prediction based on flexible neural tree, Applied Soft Computing", vol. 12, no. 1, **(2012)**, pp. 274-279.
[2] D. C. Park, "Structure optimization of BiLinear Recurrent Neural Networks and its application to Ethernet network traffic prediction", Information Sciences, vol. 237, **(2013)**, pp. 18-28.
[3] Y. J. Chen and J. Zhang, "Design of price mechanisms for network resource allocation via price of anarchy", Mathematical programming, vol. 131, no. 1-2, **(2012)**, pp. 333-364.
[4] D. Kliazovich, P. Bouvry and S. U. Khan, "DENS: data center energy-efficient network-aware scheduling", Cluster computing, vol. 16, no. 1, **(2013)**, pp. 65-75.

[5] M. E. Sousa-Vieira, "Using the Whittle estimator for VBR video traffic model selection in the spectral domain", Transactions on Emerging Telecommunications Technologies, vol. 26, no. 5, **(2015)**, pp. 731-748.
[6] P. Tune and D. Veitch, "Sampling vs sketching: An information theoretic comparison", The 30th IEEE International Conference on Computer Communications, Shanghai, China, **(2011)** April 10-15.
[7] D. C. Park, "Structure optimization of BiLinear Recurrent Neural Networks and its application to Ethernet network traffic prediction", Information Sciences, vol. 237, **(2013)**, pp. 18-28.
[8] P. Cortez, M. Rio, M. Rocha and P. Sousa, "Multi-scale Internet traffic forecasting using neural networks and time series methods", Expert Systems, vol. 29, no. 2, **(2012)**, pp.143-155.
[9] C. Katris and S. Daskalaki, "Comparing forecasting approaches for Internet traffic", Expert Systems with Applications, vol. 42, no. 21, **(2015)**, pp. 8172-8183.
[10] M. Oravec, M. Petras and F. Pilka, "Video Traffic Prediction Using Neural Networks", Acta Polytechnica Hungarica, vol. 5, no. 4, **(2008)**, pp. 59-78.
[11] L. Ding, J. She and S. Peng, An Integrated Prediction Model for Network Traffic based on Wavelet Transformation. Electronics and Electrical Engineering, vol. 19, no. 3, **(2013)**, pp. 57-62.
[12] K. Zhang, Z. Hu and X. T. Gan, "A Network Traffic Prediction Model Based on Quantum-Behaved Particle Swarm Optimization Algorithm and Fuzzy Wavelet Neural Network Discrete", Dynamics in Nature and Society, vol. 2016, **(2016)**, pp. 1-11.
[13] X. C. Sun, H. Y. Cui and R. P. Liu, "Modeling deterministic echo state network with loop reservoir", Journal of Zhejiang University SCIENCE C, vol.13, no. 9, **(2012)**, pp. 689-701.
[14] H. S. Wang and X. F. Yan, "Optimizing the echo state network with a binary particle swarm optimization algorithm", Knowledge-Based Systems, vol. 86, **(2015)**, pp. 182-193.
[15] S. P. Chatzis and Y. Demiris, "Echo state Gaussian process", IEEE Transactions on Neural Networks, vol. 22, no. 9, **(2011)**, pp. 1435-1445.
[16] L. Xiang, X. H. Ge and C. Liu, "A new hybrid network traffic prediction method", Global Telecommunications Conference, Miami, America, **(2010)** December 6-10.
[17] I. Omerhodzic, S. Avdakovic, A. Nuhanovic and K. Dizdarevic, "Energy Distribution of EEG Signals: EEG Signal Wavelet-Neural Network Classifier", International Journal of Biological and Life Sciences, vol.6, no.4, **(2010)**, pp.210-216.
[18] G. Samoradnitsky and M. S. Taqqu, "Stable non-Gaussian random processes: stochastic models with infinite variance", Chapman Hall, New York, USA, **(1994)**.
[19] L. Wang, Z. Wang, and S. Liu. An effective multivariate time series classification approach using echo state network and adaptive differential evolution algorithm. Expert Systems with Applications, vol. 43, **(2016)**, pp. 237-249.
[20] M. Reisslein, Video trace library. Arizona State University [online], Available: http://trace. eas. asu. edu, **(2016)**.
[21] N. Haghighat, H. Kalbkhani and M. G. Shayesteh, Variable bit rate video traffic prediction based on kernel least mean square method. IET Image Processing, vol. 9, no. 9, **(2015)**, pp. 777-794.
[22] R. R. Nigmatullin, C. Ceglie and G. Maione, "Reduced fractional modeling of 3D video streams: the FERMA approach Nonlinear Dynamics", vol. 80, no. 4, **(2015)**, pp. 1869-1882.
[23] M. Iftikhar, H. Mathkour and M. Imran, "A novel framework for G/M/1 queuing system based on scheduling-cum-polling mechanism to analyze multiple classes of self-similar and LRD traffic", Wireless Networks, vol. 22, no. 4, **(2016)**, pp. 1269-1284.

## Authors

**Xiaochuan Sun**, he received the M.S. degree from Guilin University of Electronic Technology, Guilin, China, in 2010, and the Ph.D. degree from Beijing University of Posts and Telecommunications, Beijing, China, 2013, respectively. He works at the College of Information Engineering, North China University of Science and Technology, Tangshan, and is currently a lecturer. The current research interests include complex network theory, reservoir computing, deep learning theory and swarm intelligence theory.

**Yingqi Li**, she received the B.S. and M.S. degrees in Guilin University of Electronic Technology, Guilin, China, in 2007 and 2010, respectively. She has been with the College of Electrical Engineering, North China University of Science and Technology, Tangshan, since 2010, and is currently a lecturer. The current research interests include statistical pattern recognition, evolutionary computation and neural networks.

**Minghui Zhang**, she received the B.S. degrees in electric automatization from North China University of Science and Technology, Tangshan, China, in 2013. She is currently working toward the M. S. degree in the College of Information Engineering at North China University of Science and Technology. Her current research interests include pattern recognition, complex network, neural networks, swarm intelligence.