Modeling and Detecting Stepping-Stone Intrusion

Yongzhong Zhang[1], Jianhua Yang[2], Chunming Ye[1]

College of Management, the University of Shanghai for Science and Technology
Shanghai, 200093 China

TSYS Computer Science Department, Columbus State University
4225 University Ave., Columbus, GA 31907 USA

Summary
Most network intruders launch their attacks through stepping-stones to reduce the risks of being discovered. To uncover such intrusions, one prevalent, challenging, and critical way is to compare an incoming connection with an outgoing connection to determine if a computer is used as stepping-stone. In this paper, we present four models to describe stepping-stone intrusion. We also propose the idea applying signal processing technology to stepping-stone intrusion detection. We present the preliminary results of applying correlation coefficients to detecting stepping-stone intrusion. The contribution of this paper is that we are the first to apply correlation coefficient to stepping-stone detection, and more importantly, it does not have to monitor a session for a long time to conclude if there is a stepping-stone intrusion. Applying DFT and Z-transform to stepping-stone detection is under way.

Key words:
Stepping-stone, network security, intrusion detection, modeling stepping-stone intrusion

1. Introduction

To make them safety, intruders always tend to compromise some computers first, which are called stepping-stones [1], and then launch their attacks to the victims they are interested in through the stepping-stones. We call this kind of attack stepping-stone intrusion. The technologies used to detect such kind attacks are called stepping-stone intrusion detection. And also, the technologies to prevent this kind of attacks are called stepping-stone intrusion prevention. Usually, once a stepping-stone intrusion is identified, prevention is necessary to protect the victim. Stepping-stone intrusion detection is an essential part in an intrusion detection system (IDS). In this paper, we focus on the discussion of stepping-stone intrusion detection, especially, the preliminary results of applying signal processing technologies to detect stepping-stone intrusion.

Part of this paper has been presented at the conference "Second International Symposium on Intelligent Information Technology Application", Dec. 2008, Shanghai, China. This research has been supported by Shanghai Leading Academic Discipline Project, Project Number: S30504

Manuscript received July 5, 2009
Manuscript revised July 20, 2009

There have been a number of stepping-stone detection methods proposed using signal processing technology recently. Most of them are just using the idea of correlation to find some kind of similarities between traffic of different connections. Actually, they are not connected to signal processing based correlation or statistic based correlation coefficient. A recent approach is called multi-dimensional flow correlation [2, 3], which uses multiple characteristics of a packet flow to do the correlation analysis instead of just one parameter. These characteristics include packet event times, packet inter-arrival times, inter-burst times, bytes per packet, cumulative bytes per packet, bytes per burst and periodic throughput samples. It is a good idea to use multiple characteristics. However, the approach does not prove advantages of using multiple characteristics.

There are very few approaches in literature which focus on the seemingly best characteristic for correlation – time interval of adjacent packets. Multi-scale Stepping-Stone Detection [4] is one which uses wavelet coefficient to compute the correlation coefficient based on time intervals of adjacent packets. The problem is this method only focus on the theoretic analysis without implementation. Moreover, it needs to monitor a session for a long time to collect enough packets to guarantee good performance, partly because multi-scale wavelet needs high computation cost.

Inter-Packet Delay Based Correlation approach [5] is the only one available in literature which addresses the correlation coefficient and focuses on time intervals of adjacent packets. It proposed four correlation point functions: 1) mini/max sum ratio; 2) statistical correlation; 3) normalized dot product one; 4) normalized dot product two. It found out through experiments that mini/max sum ratio has the best performance, and the statistical correlation is the worst. The promising aspect of this method is it could obtain correlation coefficient using only a few dozens of packets, and achieve relatively high performance with only a few dozen of packets. It shows the

time intervals of adjacent packets is a good choice, which makes it possible to distinguish different packet traffics based on very few collected packets, and meantime maintain high accuracy of detection.

There are still many problems needs to be explored to study the behavior of stepping-stone intrusion, and also advanced technologies needs to be developed to detect stepping-stone intrusion, especially, in recent years, more new and advanced technologies were adopted by intruders to attack other computers as of fast developing of computer, software, and the Internet. With existing tools or slight modification to some tools, intruders could easily attack us, and even evade our detection.

In this paper, we connect stepping-stone intrusion detection to signal processing technology. To do so, what we need to do first is to model stepping-stone detection problem to a signal processing one. There are bunch of signal processing technologies developed since 50 years before. Those technologies have been proved to be very stable, reliable, and successful in other computer and non-computer science areas. Secondly, we summarize those technologies which could be used in stepping-stone intrusion detection and discuss the challenges we may meet in terms of applying signal processing technologies to stepping-stone intrusion detection. Thirdly, we present some preliminary results we have obtained in applying correlation coefficient to detect stepping-stone intrusion. Finally, we conclude this paper, and discuss future work.

2. Model Stepping-stone Intrusion and Its Detection

Before we study stepping-stone intrusion detection, we must model stepping-stone intrusion first. Stepping-stone intrusion process is fundamentally a TCP/IP interactive accession. Intruders usually use OpenSSH [6] to establish a TCP/IP session to a victim through compromising computers in between. This session must be an interactive session because intruders' intension is to steal or look through some useful information from the victim side, not to crash the victim system. This process can be modeled as shown in Figure 1.

Fig. 1 A Connection chain

Suppose a user logs in at Host 1, and connects to Host *n* through Host 2, … , and Host *n-1*. Between any two consecutive hosts Host *i* and Host *i+1*, a connection $C_{i,i+1}$ is established. This connection is an outgoing connection to the Host *i*, and also an incoming connection to Host *i+1*. All the connections form a connection chain. We assume that Host *i* is the node that we could put our program to monitor the connection chain and capture the TCP/IP packets going through it. We also assume that the connection is set up by using OpenSSH. So we need to monitor one port of Host *i* that connects port 22 of Host *i+1*. If there is a packet sent from Host *i* to Host *i+1*, most probably, this packet is going to be acknowledged by Host *i+1* first, and then echoed by Host *i+1*. Suppose we call the packet sent from Host *i* a 'Send' packet, the acknowledge packet from Host *i+1* a 'Ack' packet, and the packet echoed from Host *i+1* an 'Echo' packet.

Any packet sent from Host *i* is originally sent from Host 1 which is the intruder's host for which we have no idea where it is. Similarly, any Echo packet sent from Host *i+1* is actually echoed back from Host *n* which is the victim side for which we also have no idea where it is located. Our objective is to determine if Host *i* is used as a stepping-stone. The basic idea to detect if Host *i* is used as a stepping-stone is to decide if any incoming connection of Host *i* is relayed by any outgoing connection of Host *i*. If it is relayed, Host *i* is used as a stepping-stone. Here what we must be clear is that using as a stepping-stone does not mean it is an intrusion. But it is highly suspicious that this connection chain is used for intrusion. This model is shown in Figure 2. $C_{i-1,\,i}$ is the incoming connection of Host *i*, and $C_{i,\,i+1}$ is the outgoing connection of Host *i*.

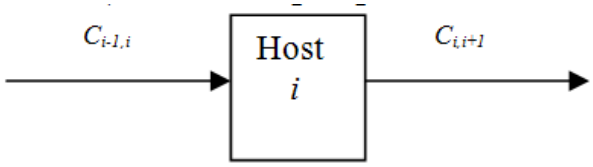

Fig. 2 Single stepping-stone model

If we could model the incoming connection of this host as one signal S_i, the outgoing connection as another signal S_o, the problem to determine if two connections are related becomes a the problem to process two signals. The question is how to abstract or model these two signals. We propose four models: Sequence Model, Pair Model, Round-trip Time (RTT) Model, and Cross Model. Different models may support different signal contents and formats. We may apply different signal processing technologies to the signals from different models. However, whatever kind of model we might use, the objective to determine if the two signals are related or correlated is kept the same.

Figure 3 shown below is used to describe these four models. We use C_{in} and C_{out} to represent any one of the incoming connections and any one of the outgoing connections of a Host *i* we are interested in, respectively. We monitor the incoming connections and outgoing connections simultaneously and collect the TCP/IP packets flowing through the connections. An incoming connection includes two streams. One is the send packet stream S_i={s_{i1}, s_{i2}, s_{i3}, …, s_{in}} from the upstream host of Host *i*, here we use timestamp received at Host *i* to represent each packet. Another one is the echo packet stream E_i = { e_{i1}, e_{i2}, e_{i3}, …, e_{im} } from the downstream host of Host *i*. For an outgoing connection, we also have two streams S_o={s_{o1}, s_{o2}, s_{o3}, …, s_{ip}}, the send packet stream, and E_i = { e_{o1}, e_{o2}, e_{o3}, …, e_{iq} }, the echo packet stream.

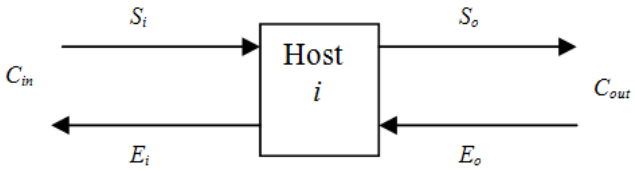

Fig. 3 Stepping-stone stream model

Model1: Sequence Model In this model, we directly take the sequence S_i and S_o as two signals. Each sequence is the collection of ordered packet timestamps. That is why it is called Sequence Model. Symmetrically, we also can consider E_i and E_o as two signals. These signals are discrete signals with time domain and each value is a timestamps. When we process these signals, there are three features we need to be aware of. One is the order of the elements, another one is the number of the elements in each sequence, and the third one is value of each timestamp.

Model2: Pair Model In Model 1, we consider send stream and echo stream separately. Actually, the two streams must be connected internally somehow as they come from the same connection. Each echo packet is the result of one or more send packets. If we take the whole part of the connection chain from Host *i* to the victim side as one box, a send packet stream is the input of the box, and the echo packet stream is the corresponding output. Here we can use the internal relation of the two streams to build a paired stream in the incoming connection, such as P_i = {(s_{i1}, e_{i1}), (s_{i2}, e_{i2})…, (s_{ik}, e_{ij}), …, (s_{in}, e_{im})}, as well as for the outgoing connection, such as P_o = {(s_{o1}, e_{o1}), (s_{o2}, e_{o2})…, (s_{ok}, e_{oj}), …, (s_{ip}, e_{iq})} , here P_i and P_o are used to represent the paired signals of the incoming connection and the outgoing connection respectively. If the two connections are relayed, the relation that any pair of P_i must be contained in one pair of P_o must be satisfied for most of the pairs. The more pairs contained, the more likely the two signals are related [7].

Model 3: RTT Model This model comes from Model 2 directly. In Model 2, we get one pair for each send packet. Each pair is actually the timestamps of the send and echo of the same packet. The difference of each pair, which is also called the round-trip time (RTT), can be used to represent the length of the connection chain from Host *i* to the victim side roughly. On the other hand, this RTT number can also be used to describe the network traffic situation at the time that packet is sent and echoed. Different pairs can provide different RTTs because the network traffic cannot always be the same, and also that the differences among the RTTs in the same sequence cannot be too big because those pairs are used to measure the length of the same connection chain. Those RTTs conform to γ-distribution [8]. We denote the RTT sequence as RTT_i = { $\Delta t_{i1}, \Delta t_{i2}, \ldots, \Delta$} for the incoming connection, and RTT_o = { $\Delta t_{o1}, \Delta t_{o2}, \ldots, \Delta t$ } for the outgoing connection. The two signals are both discrete and time domain signals. If Host *i* is used as a stepping-stone, the two signals must be related somehow. We may apply signal processing technology to process these two signals to determine if Host *i* is used as a stepping-stone.

Model 4: Crossing Model To resist intruders evasion to stepping-stone intrusion detection, Crossing Model does better than RTT Model. Crossing Model is basically a RTT Model. The difference is that instead of pair send and echo stream of the same connection we pair the send stream of the incoming connection with the echo stream of the outgoing connection, the send stream of the outgoing connection with the echo stream of the outgoing connection. They are denoted as RTT_{io} = { $\Delta t_{io1}, \Delta t_{io2}, \ldots, \Delta t$ } and RTT_{oi} = { $\Delta t_{oi1}, \Delta t_{oi2}, \ldots, \Delta t_i$}, respectively. If Host *i* is used as a stepping-stone, RTT_{io} and RTT_{oi} must have the similar distribution, and also be related somehow. Determining if a host is used as a stepping-stone can be implemented by processing these two signals with signal processing technology.

3. Signal Processing Technology Summary

Signal processing involves techniques that improve our understanding of information contained in collected data. Normally, when a signal is measured with an oscilloscope, it is viewed in the time domain. When the frequency content of the signal is of interest, it makes sense to view the signal in the frequency domain. For many signals, this is the most logical and intuitive way to view them. Simple signal processing often involves the use of gates to isolate

the signal of interest or frequency filters to smooth or reject unwanted frequencies [9, 10]. The frequency domain display shows how much of the signal's energy is present as a function of frequency. For a simple signal such as a sine wave, the frequency domain representation does not usually show us much additional information. However, with more complex signals, such as the response of a broad bandwidth transducer, the frequency domain gives a more useful view of the signal.

The technologies developed to process signals include the ones on analog signal processing and the ones on discrete signal processing. The signal processing technologies on analog signal processing include Convolution, Fourier transform, and Laplace transform [10]. The technologies on discrete signal processing mainly include Discrete Fourier Transform (DFT), Z-transform [11], and Correlation Coefficient [12]. The signals in our four models are all categorized as discrete signal. We summarize the discrete signal processing technologies in following.

DFT transforms one function in the time domain to another in the frequency domain. It requires an input function that is discrete and whose non-zero values have a limited duration. Since the input function is a finite sequence of real or complex numbers, the DFT is ideal for processing information stored in computers, such as the signals in our Model 3, and Model 4. It defines the transformation from the sequence $\{x_0, \ldots, x_{N-1}\}$ to the sequence $\{X_0, \ldots, X_{N-1}\}$ according to the formula (1),

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn} \qquad k = 0, \ldots \qquad (1)$$

,where $e^{-\frac{2\pi i}{N}}$ is a primitive N'th root of unity.

Z-transform converts a discrete time-domain signal, which is a sequence of real or complex numbers, into a complex frequency-domain representation. It was originally introduced by E. I. Jury in 1958 in Sampled-Data Control Systems. The idea contained within the Z-transform was previously known as the "generating function method". It can be defined as either a one-side shown as formula (2) or two-side transform shown as formula (3).

$$X(z) = Z\{x[n]\} = \sum_{n=0}^{\infty} x[n] z \qquad (2)$$

$$X(z) = Z\{x[n]\} = \sum_{-\infty}^{\infty} x[n] z \qquad (3)$$

,where n is an integer and z is, in general, a complex number.

In probability theory and statistics, correlation, also called correlation coefficient, indicates the strength and direction of a linear relationship between two random variables. In general statistical usage, correlation or co-relation refers to the departure of two variables from independence. Correlation Coefficients are also widely used in signal processing for analyzing similarities between two signals, either in continuous or discrete time domain. In this broad sense there are three coefficients, measuring the degree of correlation, adapted to the nature of data. These three sophisticated statistical correlation coefficients are examined based on time interval characteristic and the implementation details are in the following sections. Three coefficients, Spearman Rank (ρ), Kendall Tau Rank (, and Pearson Product-Moment (ρ), are shown in formula (4), (5), and (6), respectively.

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \qquad (4)$$

, where d_i is the difference between each rank of corresponding values of x and y, and n is the number of pairs of values.

$$\tau = \frac{4P}{n(n-1)} - 1 \qquad (5)$$

, where n is the number of items, and P is the sum of all the items ranked after the given item by both rankings.

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_x \sigma_y} = \frac{E((X-\mu_X)(Y-\mu_Y)}{\sigma_X \sigma_Y} \qquad (6)$$

, where X and Y are two random variables with expected values μ_X and μ_Y and standard deviations σ_X and σ_Y respectively, and E is the expected value operator, and cov represents covariance.

Applying DFT and Z-transform to the signals in Model 3 and Model 4 is still under way. We have made some preliminary results which are very limited. And also because of limited space, in this paper, we only present some preliminary results on applying correlation coefficient to the signals in Model 1.

4. Applying Signal Processing Technology to Stepping-Stone Intrusion Detection and Its Preliminary Experimental Results

In Model 1, Sequence Model, it has two signals with each a discrete signal. We use send packet stream as our signals. The signal for incoming connection is $S_i=\{s_{i1}, s_{i2}, s_{i3}, \ldots, s_{in+1}\}$, and for outgoing connection is $S_o=\{s_{o1}, s_{o2}, s_{o3}, \ldots, s_{om+1}\}$. The elements in these two signals only represent the timestamps of different send packets. They cannot be directly used by correlation coefficient formula (4), (5), and (6). We convert these two sequences to two discrete signals $f_1(n)$ and $f_2(m)$ by computing the intervals between s_{ij} and s_{ij+1} as the following,

$$f_1(n) = \{t_{11}, t_{12}, \ldots, t_{1n}\}$$
$$f_2(m) = \{t_{21}, t_{22}, \ldots, t_{2m}\}$$

We apply correlation coefficient formula (4) to (6) to signals $f_1(n)$ and $f_2(q)$, and get three coefficients ρ, τ, and $\rho_{X,Y}$ with each of them between 0 and 1. Each of the three coefficients can be used to determine how much the correlation is between the two signals. More close to 1, the higher correlation between the two signals. Based on our experimental results, we found that if we combine three of them to get the minimum of the three coefficients and use the minimum one to determine the relations between two signals would be better than using any one of them only in terms of keeping the false positive alarm low.

Practically we predefine a threshold ε (between 0 and 1) which is 0.15 in our experiments. We compute the difference δ between 1 and the minimum correlation coefficient and compare if $\delta <$ is satisfied. If it is, there is a high probability that the two connections are relayed which also means that the host is used as a stepping-stone. If it is not, the probability would be very low that the two connections are in a stepping-stone pair.

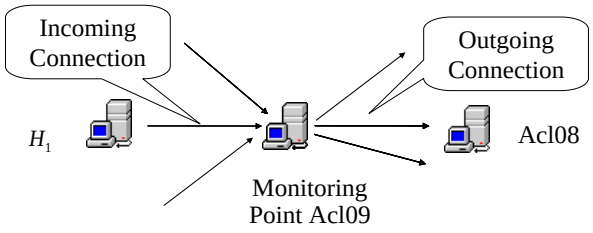

Fig. 4 Monitoring traffic in a host

We designed an experiment to verify the above idea. In the experiment, a connection chain was established via OpenSSH starting from a local host H_1 in Hampton, VA, USA, then login to another host Acl09 in National Institute of Aerospace, Hampton, VA, USA, finally from there login to a remote host Acl08 that is an assumed victim site located in Huston, Texas, USA. The connection chain is H_1 →Acl09→ Acl08 as shown in Figure 4. The local host, Acl09, is used as the monitoring host, and Acl08 is the remote victim host. We made four connections at each time experiment. All the four incoming and outgoing connections at Acl09 were monitored and recorded into files by a TCP/IP packet capturing program made with C++ language from the beginning to the end of the session.

We repeated the same experiment many times to make our results stable. All the Send and Echo packets coming in and going out the monitoring unit, local host Acl09, were monitored, collected, and recorded in a file. These data are used for analysis and computation of correlation coefficient.

Both the source H_1 and the monitoring center Acl09 were located in Hampton, VA. Outgoing connections were established between Acl09 and the destination Acl08 located in Houston, TX. In truth, the outgoing and incoming connections formed a relayed connection pair starting from H_1 and ending at Acl08. This made the monitoring sensor Acl09 a stepping-stone during the whole session. In our algorithm, we only use the send packet sequence. They are the send packets from source H_1 to stepping-stone Acl09 and the send packets from Acl09 to destination Acl08. For the stepping-stone host Acl09, it has four incoming connections, and four outgoing connections. We compute the correlation coefficients between any one of the incoming connections and all four outgoing connections. In each pair, one minimum coefficient will be picked up from three coefficients computed with the three formulas (4)-(6). We did the computations for the points (25, 55, 65, 90, 130) with each number represents the number of send packets collected. For the four pairs, we get four groups of coefficients which are shown in Figure 5. Obviously we found that the one marked with '*' has coefficients close to one while the other three pairs have coefficients much lower than one. We may conclude that this host Acl09 is used as a stepping-stone with a very high probability because we found two connections are relayed.

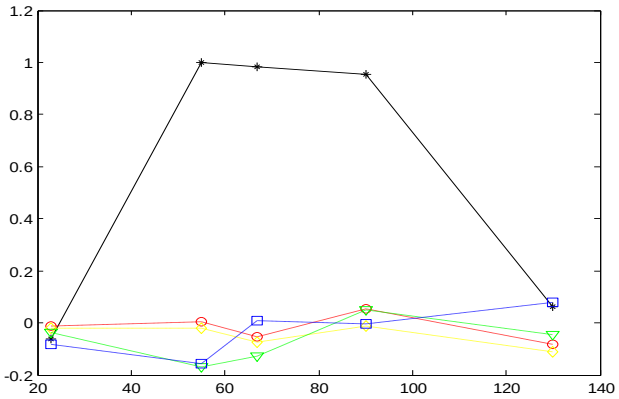

Fig. 5 Picking the minimum of Pearson, Kendall and Spearman Correlation Coefficients (black *: relayed pair, username and password only)

5. Conclusions and Future Work

In this paper we have proposed four models to describe stepping-stone intrusion, and applied correlation coefficient to detect stepping-stone intrusion. The very preliminary experimental results show that it is feasible to apply signal processing technology to stepping-stone

intrusion detection. Applying other signal processing technologies, such as DFT, Z-transform, is under way.

References

[1] Zhang, Y., and Paxson, V. (2000). "Detecting Stepping Stones". *Proc. of the 9th USENIX Security Symposium*, Denver, CO, USA, pp. 171-184.

[2] Strayer, W. T., Jones, C. E., Schwartz, B., Mikkelson, J., and Livadas, C. (2005). "Architecture for Multi-Stage Network Attack Trace back". *Proc. of the First IEEE LCN Workshop on Network Security*, Sydney, Australia, pp.15-17.

[3] Strayer, W. T., Christine Jones, Beverly Schwartz, Sarah Edwards, Walter Milliken, Alden Jackson (2007). "Efficient Multi-Dimensional Flow Correlation". *In Proc. 32nd IEEE Conference on Local Computer Networks*, LCN (2007), pp. 531-538.

[4] Donoho, D. L., et al. (2002). "Detecting Pairs of Jittered Interactive Streams by Exploiting Maximum Tolerable Delay". *Proc. of International Symposium on Recent Advances in Intrusion Detection*, Zurich, Switzerland, pp. 45-59.

[5] Xinyuan Wang, Douglas S. Reeves, S. Felix Wu (2002). "Inter-Packet Delay Based Correlation for Tracing Encrypted Connections Through Stepping Stones". *The Proc. 7th European Symposium on Research in Computer Security*, Zurich, Switzerland, pp. 244-263.

[6] Ylonen, T. (2004). "SSH Transport Layer Protocol". *Draft IETF document, http://www.ietf.org/internet-drafts/ draft-ietf-secsh-transport-18.txt*, Accessed at June 2004.

[7] Jianhua Yang, Shou-Hsuan Stephen Huang (2005). "Correlating Temporal Thumbprint for Tracing Intruders". *Proceedings of 3rd International Conference on Computing, Communications and Control Technologies*, Austin, Texas, July 2005, pp. 236-241.

[8] Jianhua Yang, Shou-Hsuan Stephen Huang, Ming D. Wan (2006). "A Clustering-Partitioning Algorithm to Find TCP Packet Round-Trip Time for Intrusion Detection". *Proceedings of 20th IEEE International Conference on Advanced Information Networking and Applications (AINA 2006)*, Vienna, Austria, April 2006, Vol. 1, pp. 231-236.

[9] Brigham, E. Oran. *The fast Fourier transform and its applications*. Englewood Cliffs, N.J.: Prentice Hall. ISBN 0-13-307505-2, 1988.

[10] Oppenheim, A. V., Schafer, R. W., and Buck, J. R. *Discrete-time signal processing*. Upper Saddle River, N. J.: Prentice Hall. ISBN 0-13-754920-2, 1999.

[11] Eliaha Ibrahim Jury. *Theory and Application of the Z-Transform Method*. Krieger Pub Co, ISBN 0-88275-122-0, 1973

[12] Hoben Thomas. *Distributions of Correlation Coefficients*. Springer, ISBN 0-38796-863-6, 1989.

Dr Yongzhong Zhang is Associate Professor and Chair of Computer Science Department at Shanghai TV University, Shanghai, China. His research interests are computer, network, and information security, and distance education. He is currently a member of IEEE. Dr. Zhang can be reached at *yzhang@shtvu.edu.cn.*

Dr. Jianhua Yang is currently Associate Professor in the TSYS Computer Science Department at Columbus State University (CSU), Columbus, GA, USA. His research interests are computer, network and information security. Dr. Yang earned his Ph.D. in Computer Science at University of Houston, TX. Before joining in CSU, Dr. Yang was Assistant Professor at University of Maryland Eastern Shore from 2008 to 2009, Bennett College for Women from 2006 to 2008, and Associate Professor at Beijing Institute of Petro-Chemical Technology, Beijing, China from 1990 to 2002. He is currently a member of IEEE. Dr. Yang can be reached at *jhyang302@yahoo.com*

Dr. Chunming Ye is Professor in the College of Management, University of Shanghai for Science and Technology, Shanghai, China. His research interests are operations management, industry engineering, planning and controlling in enterprises manufacture, manufacture scheduling, enterprises resource plan, supply chain management, and enterprises informatization. Dr. Ye has published more than 60 research papers. He is a member of Chinese Mechanical Engineering Society . Dr. Ye can be reached at *ychunming@shtvu.edu.cn.*