An Improved Integrated Prediction Method of Cyber Security Situation Based on Spatial-time Analysis

Zhijie Fan[1,2], Zhiping Tan[3], Chengxiang Tan[1], Xin Li[4]

[1] Electronics and Information Engineering School, Tongji University, China
[2] The Third Research Institute of Ministry of Public Security, China
[3] Huawei Technologies Co. Ltd, China
[4] College of Information Technology and Cyber Security, People's Public Security University of China, China
1310898@tongji.edu.cn, aaronzptan@gmail.com, jerrytan@tongji.edu.cn, lixin1999@126.com

Abstract

Cyber security situation awareness, as an effective supplement in cyber security protection measures, has been one of the research focus in recent years. In particular, cyber security situation prediction has become a hotspot of research. However, the existing cyber security situation prediction methods neglect the influence of future security elements when measuring the future security situation. Another fact is that the relationships among the security elements are always ignored. In this work, we presented an improved integrated cyber security situation prediction method based on spatial-time analysis from a new perspective. We described cyber security elements in different levels by a hierarchical index system. Then we predicted the future security elements independently in time dimension. In the process of spatial dimension prediction, we made a fusion prediction of the future security elements by using Fuzzy Cognitive Maps (FCM), and meanwhile, we corrected the prediction in spatial dimension prediction by using threat intelligence data. Finally, we used DARPA2000 datasets that is from Lincoln Laboratory Scenario (DDOS) to verify and analyze our method. The experimental result shows that the proposed method can model the future cyber security situation in network environment in a more accurate way by comparing with other similar methods.

Keywords: Cyber security, Situation prediction, Fuzzy cognitive maps, Time and spatial dimension

1 Introduction

With the rapid development of the internet, a huge number of cyber security events are springing up. The traditional cyber security protection methods. Situation Awareness (SA) is applied in the military area at the first stage. Endsley [1] divided SA into three parts: detection, understanding and prediction. Bass [2] firstly introduced SA in the area of network security, and proposed a concept named network security situation awareness. He also divided network situation awareness into three steps: extraction, assessment and prediction. The situation extraction is to obtain the basic information and some important factors related to the network security in network environment. This step completes a preprocessing for security data by real time data collection from multi-source secure data sensor. The situation assessment mainly analyzes the collected data and provide a quantificational value for network security status. The prediction is to predict the future security condition of network environment, and it is the main objective of the network situation awareness.

In this paper we study the cyber security situation prediction. There have been many researches in the field of cyber security situation prediction. However there still are several problems. For example, during the prediction stage, the security factors change with time, so the traditionally considered history data and current data, as well as the future security factors situation is fully taken into consideration. Moreover, most of traditional methods lack the analysis of common interrelation and restrictive correlation among security situation elements. Therefore, to solve the above two problems, this paper proposes an improved integrated prediction method based on space-time analysis a new perspective and gives an calculation method for cyber security situation prediction.

The rest of this paper is organized as follows. In Section 2, we overview the related work. Section 3 describes the foundation of cyber security situation prediction we proposed. Section 4 presents the proposed detailed prediction method. Section 5 shows the simulations that describe experimental analysis. Finally, concluding remarks are made in Section 6.

2 Related Works

There is a large amount of research in the area of

*Corresponding Author: Zhijie Fan; E-mail: 1310898@tongji.edu.cn
DOI: 10.3966/160792642018111906015

cyber security situation prediction. It purposes to provide an informational reference to the human analyst to help them in formulating and implementing timely preventive measures before the network is under attack. At present, there are several different techniques to complete the prediction process, the techniques contains Markov model, Bayesian network, Grey theory, Evidence theory, Neural network, Game theory etc.

Markov model consists of a list of the possible states of system, the possible transition paths between those states and the probabilities of those transitions. Recent years, Markov Model has been used to improve the cyber security situation prediction models [3-5]. HMM was used in intrusion detection system, and built the model based on alerts in order to predict cyber security situation by computing the state transition probabilities [6-7].

Bayesian network is represented by a set of radom variables and their conditional dependencies via a derected acyclic gragh. It used to complete evaluation of DDoS attacks and defense in a typical enterprise network [8-9]. A framwork was proposed to process the generated alerts in real time, correlate the alerts, and the authors constructed the attack scenarios for predicting the next goal of attackers [10]. A probabilistic methodology was designed to make inferences in abnormal situations. It was trained to understand human interactions and crowd behavior to complete prediction [11].

Evidence theory, especially Dempster-Shafer (D-S) theory is a general framework for reasoning with uncertainty. It was used to predict cyber security situation. The description of uncertainty is added to avoid absolute predict [12], and was designed a cyber security situation perception system named Net-SSA with D-S evidence theory [13]. D-S theory has a strong ability of fusion multi-source cyber scurity information from different network equipment. The prediction of cyber security situation considered multi-source information, it was used for prediting combied other methods [15-16].

Grey theory was first proposed in [17]. It is used to predict from the grey system which lack of information. It is applied to build a dynamic model with a group of differential equations, which is called Grey Model (GM). The model is able to prevail over the weakness of probability and discovered the relationship among the limited and confused data. It can weaken the randomness of the original sequence. This makes the process to find out the variation regularity in the sequence easier and use this regularity to predict [18-21].

Neural network represents a computing method that simulates the way that the brain performs computations. It was used to completet cyber security situation prediction [22]. The impact of sample covariance and noise on the network training is considered and the traditional function of error is replaced by the maximum likelihood error function. Through the error analysis, the predicted error value will be obtained and feedback to the prediction model as the training signal for the situation index weights adjustment [23-25].

Game theory is the study of mathematical models of conflict and cooperation between intelligent rational decision-makers [26]. It compositively analysed influence of network behavior by considering three factors that are attack behavior, defense behavior and ordinary users behavior. It also provided the best reinforcement scheme according to the action space of the attacking party. But, with the increasing of network scale, it can not provide continuing accurate analysis, and the high load led to bad extendibility [27].

In recent years, various other different methods in the field of cyber security situation prediction have been proposed. The authors proposed an improved prediction model named SCGM(1,1)c model that can be used for predicting time series of cyber security situation [28]. A method was came up with a cyber security situation prediction method based on spatial data mining analysis and time sequence analysis, however, it only makes a simple superposition of time and space [14]. Another system was developed by using attack graph model, it used Attack Graph (AG) and service dependencies to describe cyber security situation values over time [29]. Fuzzy Cognitive Maps (FCM) was first used to analyze risk impact factors to assess cyber security situation, but it has not been used in the field of cyber security situation prediction [30]. A pattern-matching method has been proposed to provide the basic seurity information for network security situation prediction [31]. We compared typical cyber security situation prediction methods in Table 1.

The above research solved the problems of predicting the cyber security situation from different perspectives. In this paper, we propose an improved integrated prediction method based on spatial-time analysis and describe method for calculating cyber security situation prediction. Our method considers the impact of both history and current data on the future cyber security situation, and also considers the impact of future data. Our method reflects the common interrelation and restrictive correlation among the different security factors of the cyber security situation.

3 Foundation of Cyber Security Situation Prediction

3.1 Basic Concepts

In order to better and clearly describe the method of cyber security situation prediction, the related terminologies are defined as follows.

Table 1. Comparison of different typical cyber security situation prediction methods

	Method Category	Modeling Time	Model Space	Prediction Time	Knowledge Source	Number of Feature	Scalability
Mathematics Formula	mathematics model	short	small	short	experience	small	hard
Markov Model	knowledge reasoning	medium	big	medium	expert, probability	medium	medium
Bayesian Network	knowledge reasoning	medium	big	medium	expert, probability	medium	medium
Gray Theory	pattern recognition	medium	medium	long	experience, gary relation degree	big	medium
Evidence Theory	knowledge reasoning	medium	medium	long	probability, evidence combination	medium	medium
Neural Network	knowledge reasoning	long	medium	medium	machine learning	medium	easy
Game Theory	knowledge reasoning	medium	big	medium	expert, probability	medium	hard

Definition 1. A Cyber Security Situation denoted by E is defined as the current security status and future development trend in the target network.

Definition 2. f is a function of $E_e(i)$, where $E_e(i)$ is a set of cyber security situation elements that is defined as the basic data elements.

Cyber Security Situation Elements are used for calculating cyber security situation. The extracting of situation factors is the first step of situation prediction. By calculating the factors status we can get the overall cyber security situation, i.e., $E=f(E_e(i))$.

Definition 3. T is a Cyber Security Situation Index System that T is used to describe and quantify the cyber security situation.

Definition 4. $D_L=\{D_{Lh}, D_{Lc}, D_{Lf}\}$ is a Local Data which is the cyber security related data in target network system during prediction, where D_{Lh}, D_{Lc}, and D_{Lf} are a set of history data, current data, and future data, respectively.

Definition 5. Intelligence Data D_I is a set from different source network which is similar to the target network of cyber situation prediction, including similar cyber security data, facility security data, software and business security data, etc.

Definition 6. $D_{Lf}(i)=f(D_{Lh}(i), D_{Lc}(i))$ is defined as a time dimension prediction, where $D_{Lh}(i)$ and $D_{Lc}(i)$ are history and current data of the i-th situation element, respectively.

In the above definition, f is used for the prediction of the time dimension. As for a certain situation factor which affects the cyber security, we can obtain the future data $D_{Lf}(i)$ according to the history and current data.

Definition 7. $D_S(j)=g(D_{Lf}(i), D_I(i))$ is defined as a spatial dimension prediction, g is used for the prediction of the spatial dimension.

Note that the value of the spatial dimension prediction function depends on two parts, one is the future data $D_{Lf}(i)$ of situation factors according to the interrelationship among different factors and another is intelligence data $D_I(i)$.

The computing of cyber security elements $E_e(i)$ is expressed as $E_e(i)=f(D_{Lh}(i), D_{Lc}(i), D_{Lf}(i), D_I(i), T)$ in this work. As shown in Figure 1, the integrated prediction method can be divided into prediction of time and space, as well as modification with intelligence data. Firstly, we independently predict the situation factors in time dimension, and then predict that in spatial dimension. The spatial dimension prediction includes predicting the future data of situation factors according to the interrelationship among different factors and correcting the prediction by using intelligence data. After that, the status of future cyber security situation factors can be calculated, and the cyber security situation also can be obtained.

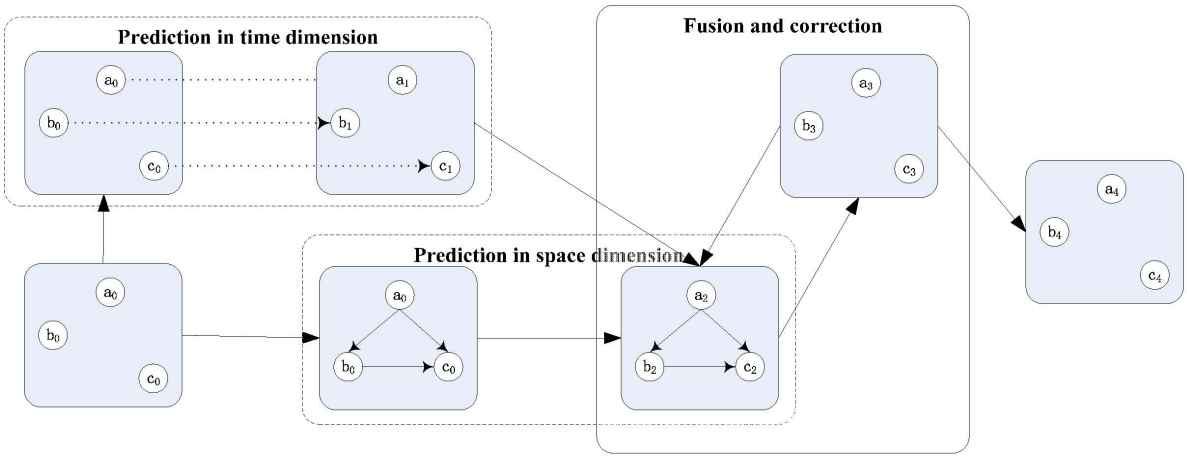


Figure 1. Cyber security situation prediction method

3.2 Hierarchical Quantificational Index System

The hierarchical quantificational security index system was defined firstly in [32], it is used to describe cyber security situation in this work. As from Figure 2, the cyber security index set is divided into three index layers. The first layer is used to describe the Cyber Security Situation Integrated Index (CSSII). The index in the first layer is decided by several indexes from the second layers, including Infrastructure Security Index (ISI), System Vulnerability Index (SVI), System Threat Index (STI), etc. In the same way, indexes from the second layer are decided by several indexes from the third layers. For example, ISI is decided by Network Traffic Index (NTI), Service Status Index (SSI), Resource Consumption Index (RCI), etc.

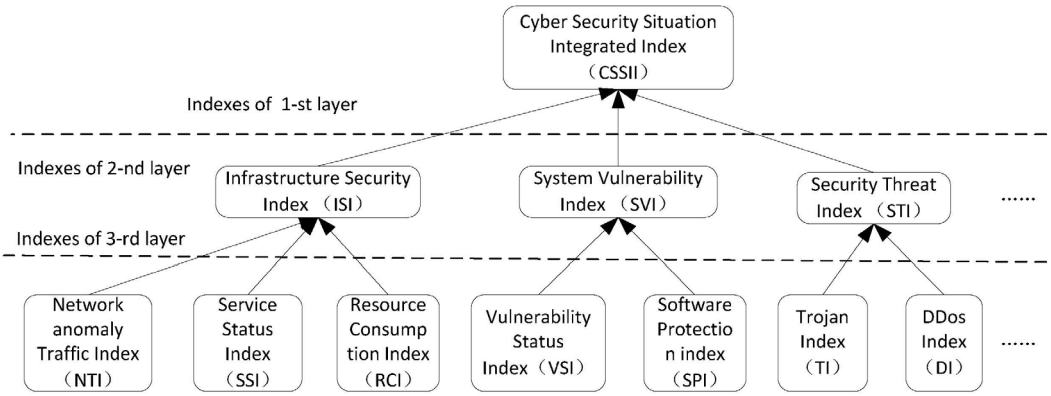


Figure 2. Hierarchical quantificational index system

In this paper, we use the third index layers as the factors in cyber security situation, so the key point of prediction method is about how to accurately predict the status values indexes of the third layer. Several prediction methods for the indexes of the third layers based on time sequence analysis have been elaborately and independently proposed in relevant literatures. In this paper, we take VSI as an example to explain how to predict the VSI based on time dimension. The following formula is used to calculate the first and second layer indexes in hierarchical quantificational index system.

$$E(t) = f\left(\sum_{i=1}^{m} e_i(t) w_i\right) \quad \textbf{(1)}$$

where $E(t)$ is the index status values on first and second layer, $f(x)$ is a normalization function, $e_i(t)$ is the status value to decide second and third layer index, and w_i is the corresponding weight.

3.3 Situation Prediction in Time Dimension

For every predictable third layer indexes, time dimension prediction is an independent single element prediction. The status values at time $t+1$ can be calculated out based on time dimension prediction according to the status values at time t.

Suppose that there are n predictable indexes of third layer in all m indexes of third layer, denoted by Et_i, $(i=1, 2, \cdots, n)$, where $n \leqslant m$, the value of $Et_i(t)$ is used to compute the value of $Et_i(t+1)$ according to $Et_i(t+1)=f(Et_i(t))$. $Et_i(t)$ and $Et_i(t+1)$ are the index status values of third layer at time t and $t+1$, respectively, and f is particular computing method that is discussed in the following content.

We take VSI as an example to introduce how to calculate $Ev_i(t+1)$ from $Ev_i(t)$ based on time dimension prediction. Vulnerability of system software is the system deficiency. It is one of the most important factors to cause system frangibility. It can cause program exception, system breakdown and become an injection entrance for hacker. In the area of vulnerability prediction, many models have been raised. In [33], the authors proposed a prediction vulnerability model based on the type of vulnerability severity. SVM was used to make an construction of a categorization framework for CVE based on SVM, which has classification ability in CVE dictionary [34]. The AML model that has been widely used, but the model lacks the prediction of multi-growth circle. Then, they raised a prediction method of multi-growth circle, but the method treats all the vulnerabilities equally [35-36]. In [14], a new method was put forward, the vulnerability prediction method is based on privilege classification, however the method do not consider the importance of the different vulnerabilities. In [29], the authors improved the vulnerability prediction accuracy with secure coding standard violation measures.

In this work, we propose a new improved vulnerability prediction method, in which we add the description of influence degree about different vulnerabilities. We divide vulnerability into eight types, *Cpphyaccess, Cpaccess, Cpsubuser, Cpuser, Cpsubouser, Cpouser, Cpsubroot* and *Cproot* according to [36]. The formula to get $EV(t)$ is as follows.

$$EV(t) = g\left(\sum_{i=1}^{n} CVS_i(t) \times w_i\right) \quad \textbf{(2)}$$

where g is normalized function, w_i is weight value of all types vulnerability, and $CVS_i(t)$ is the i-th vulnerability set in time t that is calculated as follows.

$$CVS_i(t) = f(vn(t), vpt_i(t), vf_i(t-1)) \quad (3)$$

where $vn(t)$ is the total number of vulnerabilities at time t, and $vpt_i(t)$ is the latest publish time of i-th type vulnerability at time t. $vf_i(t-1)$ is occurrence probability of i-th type vulnerability in time $t-1$. The computing process from $EV(t)$ to $EV(t+1)$ is described by using the following steps.

(1) Divide vulnerabilities into n types according to [30].

(2) Use HVS and $CVS_i(t)$ to calculate $vn(t+1)$ and $vf_i(t)$ at time $t+1$ based on the AML method [35], where HVS is history vulnerabilities set.

(3) Obtain the $vf_i(t)$ using following formula.

$$vf_i(t) = ftp_i(t) \Big/ \sum_{i=1}^{n} ftp_i(t) + \varphi \quad (4)$$

where $ftp_i(t)$ is occurrence time of i-th type vulnerability at time t, φ is a constant that is gave by actual experiment and investigation.

(4) Compute respectively the n types current vulnerabilities $CVS_i(t+1)$ by using equation (3).

(5) Calculate the weight values w_i of different types of vulnerabilities based on the severity in CVE library, and then compute the $EV(t+1)$ values of VSI at time $t+1$ according to equation (2).

For other kinds of predicable indexes of third layer, the index status values at time $t+1$ can be calculated by values at time t by using some particular methods.

3.4 Situation Prediction in Spatial Dimension

The traditional processes of cyber security situation prediction always lack of analysis of interrelations and restrictive correlations among different security factors, which usually exist in reality. In this article, the spatial dimensional situation prediction uses Fuzzy Cognitive Maps (FCM) to model a hierarchical three-layer indexes system.

Fuzzy Cognitive Maps (FCM) is firstly proposed by professor Kosko [37], who combines cognitive map with fuzzy set theory. FCM model consists of nodes, directed arcs and weights of directed edges. It is a weighted directed graph that can describe causal relationship. The node in FCM is called concept node that can describe the abstract things, concrete things, activities, system properties and system statuses according to actual demand. The weighted directed edges in FCM structure are used to describe the causal relationship between any two concept nodes. Directed edges can be viewed as single layer neural network with feedbacks and the object-oriented concept. The knowledge is inside of concept nodes and weighted directed edges. FCM model uses weighted directed relationships to simulate fuzzy reasoning, in which the interrelationships are used to stimulate dynamic behavior of system.

Definition of FCM: all concept nodes $c_1, c_2, \cdots, c_i, \cdots, c_n$ exist in FCM, the value's range of weighted directed edges is [-1, 1], e_{ij} is weight value of edge $<C_i, C_j>$, the matrix $E=f(e_{ij})$ is called an adjacent matrix or incidence matrix of FCM.

FCM can stimulate the operation states of system. The evolution process of FCM model includes forward and backward evolution. The forward evolution is mainly used for decision support and prediction, and backward evolution mainly is used for reason trace. In this article, we used the forward evolution to predict the cyber security situation. After building the FCM model and obtaining the initial status values of all concept nodes, the status values of all concept nodes at any time can be calculated according to forward evolution by the following formula.

$$A_i(t+1) = f\left(A_i(t) + \sum_{j=1, j\neq i}^{n} A_j(t) \times w_i \right)$$

Suppose that $C=\{ c_1, c_2, \cdots, c_i, \cdots, c_n \}$ is a set of all concept nodes, $n=|C|$, and C_i is the value of the i-th concept node, recorded as A_i after mapping to range [0, 1], and it means the status value of concept node. $A_i(t)$ means the status value of i-th concept node at time t, and $A_i(t+1)$ means the status value of i-th concept node at time $t+1$. w_{ji} is the incidence matrix of concept nodes, also named as adjacent matrix. f is a function, the two or three valued step function and S-curve function are commonly used in practice.

Building an FCM model consists of several steps: selecting the concept nodes, connecting the causal relationship between any two concept nodes, and determining the impact degree of causal relationship. In this work, we choose all the indexes of third layer as the concept nodes in FCM, the causal relationship and adjacent matrix can be decided respectively by typical machine learning technology. In addition, we use the forward evolution procedure to predict cyber security situation and introduce threat intelligence data in similar network system to modify situation prediction.

4 Integrated Situation Prediction Method In Cyber Security

For the hierarchical cyber security situation index system, independent prediction on single factors for every predictable indexes of third layer in time dimensional is in order to get the status of future situation in time dimensional. At the same time, we model every indexes of third layer by using FCM and introduce threat intelligence data in similar network system to modify situation prediction. The calculating method of third layer index status value at time $t+1$ can be expressed as follows.

$$e_i(t+1) = f\left(e_i(t) + \sum_{i=1, j\neq i}^{n} e_i(t) w_{j,i} + g(e_i(t)) + \theta \right) \quad \textbf{(5)}$$

where $e_i(t)$ is the status value of the i-th index of third layer at time t, w_{ji} is adjacent matrix in FCM model. $g(e_i(t))$ means status value at time $t+1$ computed via particular method from the i-th predicable index of third layer, where g is the computing method in time dimension. For the indexes that are not predicable, we set $g(e_i(t))=0$. θ is corrected parameter, and f is S-curve function.

Equation (5) is the key computing method in our prediction model, and every parts in equation (5) has reality value. The first part $e_i(t)$ is the status value of the i-th index of third layer at time t, which is construed as history data. The second part is the changeable value of i-th index of third layer in spatial dimension, which is regarded as current data. The third part $g(e_i(t))$ is the state value for i-th index of third layer at time t based on time dimension, which is regarded as future data. θ is thread intelligence data in similar network system, which is regarded as data in parallel space. In this formula, the adjacent matrix w_{ji} can describe the influence relation and degree between any two indexes of third layer. The method in this paper can solve the two above problems in the current cyber security situation prediction research.

While the computing of cyber security situation is completed, we introduce cyber security state level table defined as in [38] to reflect quantitatively the cyber security situation. The security state level is divided into excellent, fine, middle, poor and danger with every level corresponding to a range, [0, 0.2], [0.2, 0.4], [0.4, 0.75], [0.75, 0.9] and [0.9, 1], respectively, and with the corresponding weight values are 0.06, 0.11, 0.21, 0.26 and 0.36, respectively.

Suppose that the index value of third layer is $e_i(t)$, the computing procedure of CSSII value $E(t+1)$ in cyber security situation at time $t+1$ is as follows.

(1) Divide all indexes of third layer into two types, one is predictable in time dimension denoted by $e'_i(t)$, and the anther is not predictable denoted by $e''_i(t)$.

(2) Choose a suitable algorithm or function g for predicable indexes to calculate the values of $g(e_i(t))$.

(3) Use equation (5) to compute $e'_i(t+1)$, and during this procedure, if the index is not predicable, set $g(e_i(t))=0$.

(4) Combine $e'_i(t+1)$ and $e''_i(t+1)$ into a new set $e_i(t+1)$, and use equation (2) to get the index values of second layer $d_i(t+1)$.

(5) Use equation (1) to get the index value of first layer $E(t+1)$.

(6) Give out the situation security level at time $t+1$ by contrasting the value $E(t+1)$ with the cyber security status level table.

In this work, we put forward a prediction method from a new perspective. We put the independent single prediction in time dimension into the iterative computations of FCM model, and use thread intelligence data to correct them in order to avoid absolute prediction. In such way, the prediction can conform to the expected situation result. The detailed prediction process is described in the following experiment analysis.

5 Experimental Analysis

This section records the analysis of experiments by using DARPA2000 Data Set from MIT Lincoln Laboratory [39]. DARPA2000 Data Set is regarded as a standard data set in cyber security field which includes comprehensive data and instructions documents.

5.1 Instruction of Experimental Environment

DARPA2000 Data Set has two DDos attack scenes which is LLDOS 1.0 and LLDOS 2.0.2, respectively. In our experiment, we chose LLDOS 1.0 as the target network system to predict and analyze the status of cyber security situation, and took LLDOS 2.0.2 as the source of threat intelligence data of similar network system. The topological structure diagram of LLDOS 1.0's attack scene is shown in Figure 3. The vulnerability information in our experimental system is shown in Table 2.

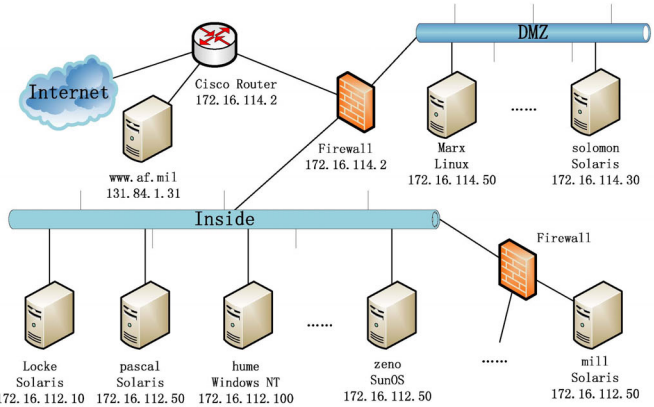


Figure 3. Network topology of DARPA2000 environment

Table 2. Vulnerability information in experimental system

No.	Host	Vulnerability	Weight	Type
1	mill, locke, pascal, hume, zeno	ICMP Incorrectly Configured	0.1	Cpphyaccess
2	mill, locke, pascal	SunRPC Incorrectly Configured	0.2	Cpphyaccess
3	mill, locke, pascal	Sadmind Buffer Overflow (CVE-1999-0977)	0.8	Cpaccess
4	mill, locke, pascal	RCP Incorrectly Configured	0.2	Cpphyaccess
5	mill	HINFO Query Incorrectly Configured	0.6	Cpphyaccess
6	www.af.mil	Syn Flood (CVE-1999-0116)	1.0	Cproot

LLDOS 1.0 is generated by a real multi-step attack which includes five steps. The first step is scanning the IP addresses form the network and attempting to find live hosts. The second step is checking all the live hosts and discovering the host which opened the *sadmind* service. The third step is launching a buffer overflow attack based on *Sadmind Buffer Overflow* bug to the hosts (*Locke*, *Pascal* & *mill*) which opened the *sadmind* service in order to get the permission of executing program on these hosts. The forth step is installing *mstream* program to the hosts. The fifth step is launching a DDOS attack based on *Syn Flood* bug to *www.af.mil* host by remotely operating the hosts which was installed *mstream* program.

In this experiment, we selected typical indexes as in Table 3. The experiment was carried out based on the five above attack steps. The cyber security situation of the target network system was divided into six statuses which are the initial state and five attack states, respectively. The initial state is the state before the attack happens, which is recorded as t_0. The five attack states are recorded as t_1, t_2, t_3, t_4, and t_5 respectively. With the attack progressing, we used our methods and the index values of third layer at t_0 to predict the index values of third layer at t_1, t_2, t_3, t_4, and t_5, and then worked out the corresponding values of the primary indexes. Finally, we compared the prediction result with ARMA method and the method in [14].

5.2 The Prediction in Time Dimension

The prediction of time dimension is a single prediction about the values of the predictable indexes of third layer in time dimension. The experiment gave an independent prediction with eight index values of third layer as shown in Table 3. Now, The instruction of the predictive test about VSI by adopting our method that is proposed in section 3.3 is mentioned in the following content. In our experiment, the computing process of the initial VSI is as follow.

Table 3. Index set of cyber security situation

Indexes of 1-st layer	Indexes of 2-nd layer	Indexes of 3-rd layer	No.
Cyber Security Situation Integrated Index (CSSII)	Infrastructure Security Index (ISI)	Network anomaly Traffic Index (NTI)	1
		Service Status Index (SSI)	2
		Resource Consumption Index (RCI)	3
	System Vulnerability Index (SVI)	Vulnerability Status Index (VSI)	4
		Software Protection index (SPI)	5
	Security Threat Index (STI)	Trojan Index (TI)	6
		DDos Index (DI)	7

(1) Classified the types of vulnerabilities according to the vulnerable information in Table 3. Our experiment system had three types of vulnerabilities: *Cpphyaccess*, *Cpaccess* and *Cproot*. Then we calculated the type weights of the vulnerabilities according to Table 2. So we obtained $w_{Cpphyaccess}=0.22$, $w_{Cpaccess}=0.34$, and $w_{Cproot}=0.44$.

(2) Selected AML model to calculate the amount and delivery time of vulnerabilities according to equation $y(t_0)=1/(e^0+1)$.

(3) Set the experiment system in one-second interval, which means that the five steps will start in every one-second respectively. According to the above equation, the amount of vulnerabilities is 0.5 at time t_0, i.e., $vn(t_0)=0.5$.

(4) Calculated the adjustment values of every types of vulnerabilities according to their occurrence number on hosts, and we obtained $\varphi_{Cpphyaccess}=5/9$, $\varphi_{Cpaccess}=3/9$, and $\varphi_{Cproot}=1/9$.

(5) Calculated the all probabilities of three different types of vulnerabilities that happening at different time respectively according to equation (4), and we obtained $vf_{Cpphyaccess}(t_0)=0+\varphi_{Cpphyaccess}=0.56$, $vf_{Cpaccess}(t_0)=0.33$, and $vf_{Cproot}(t_0)=0.11$.

(6) Calculated the vulnerabilities set $CVS_i(t_0)$ at different time of three different vulnerabilities types at time t_0 according to the equation $CVS_i(t_0)=e^{vn(t0)*vfi(t0)}$, and we obtained $CVS_{Cpphyaccess}(t_0)=1.32$, $CVS_{Cpaccess}(t_0)=1.18$, and $CVS_{Cproot}(t_0)=1.05$.

(7) Calculated $EV(t_0)=0.11$ according to equation (2).

Next is the computing process of prediction values $EV(t_1)$ by using $EV(t_0)$ of VSI after the first attack step.

(1) Selected the AML model to calculate the amount and delivery time of vulnerabilities at time t_1. According to the equation $y(t_1)=1/(e^{-(t1-t0)}+1)$, the amount of vulnerability is 0.72, i.e., $vn(t_1)=0.72$.

(2) Calculated the adjustment values of every type of vulnerabilities according to their occurrence number on hosts, and we obtained $\varphi_{Cpphyaccess}=5/9$. In the similar way, $\varphi_{Cpaccess}=3/9$ and $\varphi_{Cproot}=1/9$.

(3) Calculated all the probabilities of three different types of vulnerabilities happening at time t_1 respectively according to equation (4), and we obtained $vf_{Cpphyaccess}(t_1) = 5/16+\varphi_{Cpphyaccess}=0.87$, $vf_{Cpaccess}(t_1) = 3/16+3/9=0.52$, and $vf_{Cproot}(t_1) = 1/16+1/9=0.17$.

(4) Calculated the vulnerabilities set $CVS_i(t_1)$ at time t_1 of three different types of vulnerabilities according to $CVS_i(t_1)=e^{vn(t1)*vfi(t1)}$, and we obtained $CVS_{Cpphyaccess}(t_1) =1.87$, $CVS_{Cpaccess}(t_1)=1.45$, and $CVS_{Cproot}(t_1)=1.13$.

(5) Calculated $EV(t_1)=0.13$ according to equation (2).

In the similar way, we worked out the VSI at t_2, t_3, t_4, and t_5 as shown in Table 4.

Table 4. Prediction values of VSI in time dimension

	vn	$vf_{Cpphyaccess}$	$vf_{Cpaccess}$	vf_{Cproot}	$CVS_{Cpphyaccess}$	$CVS_{Cpaccess}$	CVS_{Cproot}	EV
t_0	0.5	0.56	0.33	0.11	1.32	1.18	1.05	0.11
t_1	0.72	0.87	0.33	0.11	1.87	1.27	1.08	0.13
t_2	0.88	1.24	0.33	0.11	2.98	1.34	1.10	0.16
t_3	0.95	1.24	0.52	0.11	3.25	1.64	1.11	0.18
t_4	0.98	1.30	0.70	0.11	3.58	1.99	1.12	0.20
t_5	0.99	1.30	0.70	0.17	3.63	2.01	1.19	0.22

During the experiment process, our method focused on an integrated prediction method of cyber security state. We came up with an improved independent prediction method of predicting VSI. As for other indexes of third layer which have predictability, such as NTI, DI and so on, there have been a lot of independent prediction methods in relative literatures in single time dimension. In our experiment, we assigned 0 to the indexes status values of third layer that are not predictable. All the prediction indexes status values of third layer in time dimension are shown in Table 5.

Table 5. Prediction index values of third layer in time dimension

	NTI	SSI	RCI	VSI	SPI	TI	DI	VI
t_0	0.130	0.000	0.320	0.120	0.000	0.000	0.210	0.000
t_1	0.630	0.000	0.520	0.130	0.000	0.000	0.270	0.000
t_2	0.450	0.000	0.470	0.160	0.000	0.000	0.310	0.000
t_3	0.200	0.000	0.700	0.180	0.000	0.000	0.500	0.000
t_4	0.320	0.000	0.410	0.200	0.000	0.000	0.810	0.000
t_5	0.880	0.000	0.810	0.220	0.000	0.000	0.970	0.000

5.3 The Prediction in Spatial Dimension

For the prediction in spatial dimension, we modeled all the indexes of third layer based on FCM, and simulated different states of the experiment system which was under dynamically attacking, in order to get information about cyber security status. We set all the indexes of third layer as concept nodes of FCM model. The relationships between any two concept nodes and the weight values on arcs were confirmed by the learning method in [40]. The concept nodes of FCM model were selected from Table 3, and the structure of FCM model is shown in Figure 4.

Next, the computing process of predicting the indexes values of third layer at t_1 according to the index values of third layer at t_0 is as follow.

(1) Obtained the set of every index values of third layer in time dimension at t_0 according to Table 2, and we obtained $Eth(t_0)$: *{0.13, 0.00, 0.32, 0.12, 0.00, 0.00, 0.21, 0.00}*.

(2) Worked out the indexes status values of third layer by using our prediction method in time dimension, and we obtained $g(e_i(t_1))$: *{0.63, 0.00, 0.52, 0.13, 0.00, 0.00, 0.27, 0.00}*.

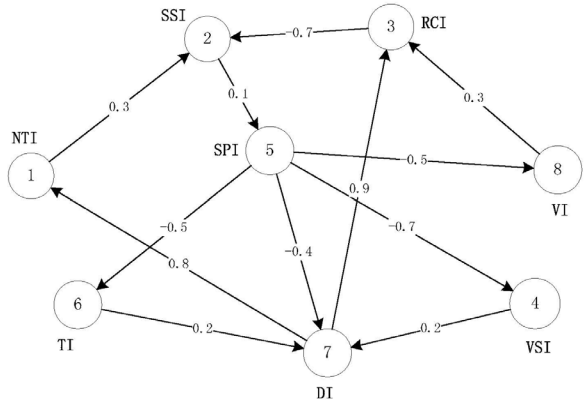


Figure 4. FCM model of third layer indexes

(3) Calculated the prediction index status values of third layer in spatial dimension according to equation (5). We selected S-curve function as transformation function: $f(x)=1/(1+e^{-cx})$, where c is the parameter *4* in our experiment. Then we worked out the indexes status values of third layer in spatial dimension at t_1: $Eth(t_1)$: *{0.976, 0.741, 0.984, 0.731, 0.5, 0.5, 0.882, 0.5}*.

(4) Introduced threat intelligence data into our experiment from similar network system. We used the difference values of negative feedback between $Eth(t_1)$ and $g(e_i(t_1))$ as parameter θ in our experiment, and we obtained the final prediction result of cyber security state by equation (5).

In similar way, we worked out every index status values of third layer at different time in spatial dimension in our experiment as shown in Table 6.

Table 6. Prediction index values of third layer in spatial dimension

	NTI	SSI	RCI	VSI	SPI	TI	DI	VI
t_0	0.130	0.000	0.320	0.120	0.000	0.000	0.210	0.000
t_1	0.776	0.541	0.784	0.531	0.300	0.300	0.682	0.300
t_2	0.846	0.699	0.832	0.741	0.509	0.531	0.893	0.531
t_3	0.904	0.848	0.974	0.844	0.793	0.792	0.997	0.792
t_4	1.000	0.817	0.905	0.908	0.707	0.609	0.999	0.609
t_5	1.000	1.000	1.000	0.960	0.987	0.728	1.000	0.728

5.4 The Prediction of Total Status of Cyber Security

In our experiment, after prediction by using our method in time and spatial dimension, we worked out the indexes state values of second layer: ISI, SVI, STI at t_0 to t_5 and the index state values of first layer according to equation (1), as shown particularly in Table 7.

Table 7. Prediction index values of first and second layer

	ISI	SVI	STI	CSSII
t_0	0.134	0.079	0.103	0.099
t_1	0.685	0.452	0.487	0.521
t_2	0.787	0.662	0.708	0.705
t_3	0.901	0.827	0.892	0.864
t_4	0.912	0.840	0.800	0.847
t_5	1.010	0.969	0.861	0.955

The trend graph of status values about indexes of second layer: ISI, SVI, and STI at t_0 to t_5 are shown in Figure 5.

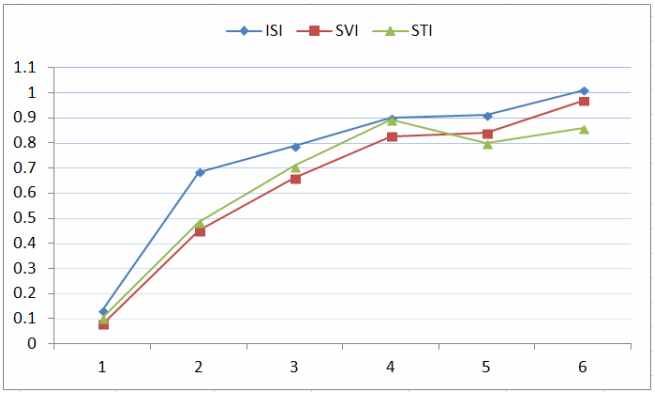


Figure 5. Trend chart of 2-nd layer indexes

Next, the computing process of CSSII $E(t_0)$ according to the prediction indexes of third layer is as follow:

(1) Worked out the index status values of second layer respectively according to equation (1). Each weight set of third layer index and the each weight set of third layer index: $w_1=\{0.27, 0.43, 0.31\}$, $w_2=\{0.66, 0.34\}$, $w_3=\{0.28, 0.49, 0.23\}$.

(2) Worked out the index status values of first layer according to equation (1) and the each weight set of second layer index: $w=\{0.26, 0.51, 0.23\}$.

5.5 Comparative Analysis

DARPA2000 Data Set has two DDos attack scenes which is LLDOS 1.0 and LLDOS 2.0.2, respectively. In experiment, we chose LLDOS 1.0 as the target network system to verify our method. We compared our method with other similar methods of predicting cyber security situation, including typical ARMA method and method in [14]. ARMA model is a typical analysis method in time sequence, which focuses on the influence on cyber security state by attack process over time. It is a typical prediction method of just in time axis. The contrast result of CSSII of the target network system at t_0 to t_5 between our method and ARMA method is shown in Figure 6, and the contrast result between our method and the method in [14] is shown in Figure 7.

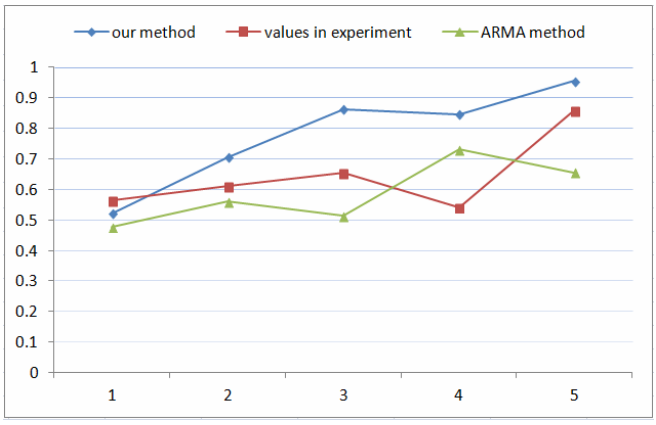


Figure 6. Comparison with ARMA method

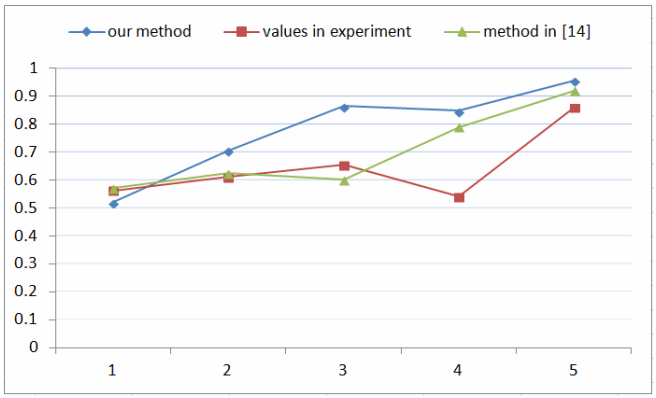


Figure 7. Comparison with method in [14]

In previous experiment process, we tested our method based on the basis of DARPA2000 database's initial vulnerability information and five given attack steps. Furthermore, in order to test and verify the effectiveness and accuracy of our method, we removed the vulnerability of *Sadmind Buffer Overflow* of *Locke* in the following experiment, which would cause the changing of the third and the forth step of attack process, and after having changed the security strategy, the contrast graph of CSSII between our method and ARMA method is shown in Figure 8 and Figure 9 shows the contrast results between our method and the method in [14].

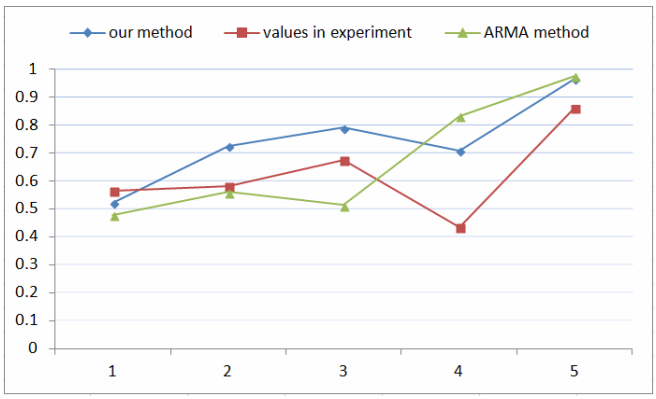


Figure 8. Comparison with ARMA method after changing configuration

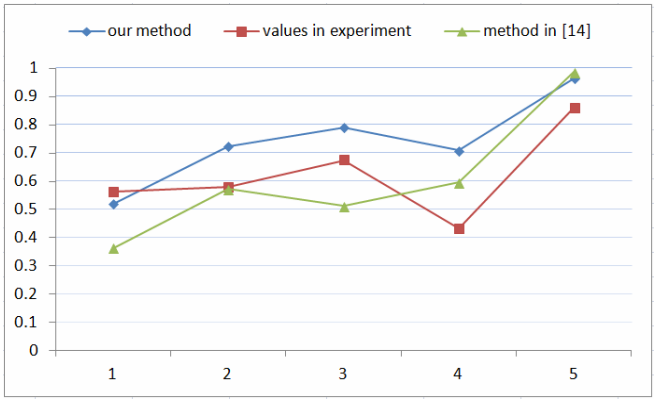


Figure 9. Comparison with method in [14] after changing configuration

Through comparing our method with other two similar methods, we can find that our method can more accurately reflect the trends of cyber security situation than ARMA method and the method in [14]. The main reason is that ARMA model is a typical analysis method in time sequence, which focuses on the influence of cyber security state during attack process, and neglects the influence of future security elements when measuring the future security situation, and it also neglects the change of rules among security elements over time. Although the method in [14] has taken consider of the influence of future security elements, it neglects the influence of the rules' change among security elements. We proposed an prediction methods based on time dimension and spatial dimension, that effectively fused the prediction method in time dimension and spatial dimension by using FCM model. Furthermore we used threat intelligence data to make adjustment of parameters. Our method in experiment simultaneously used prediction analysis in both time dimension and spatial dimension. We considered both the changing situation of all the cyber security factors in time dimension when under attacking and all the factors' state in the future which the factors determined the future's state. We also focused on the relevance among all the cyber security factors. The experiment shows that our method can more accurately reflect the trends of cyber security situation than ARMA method and the method in [14].

6 Conclusion

In this paper, we proposed a new prediction method based on spatial-time analysis from a new perspective. The method introduced spatial dimension prediction and thread intelligence data, and can avoid absolutely prediction effectively, meanwhile, the result of prediction was more accurately.

During the prediction in experiment, firstly we used hierarchical index system to describe elements of cyber security situation. Then we independently predicted each elements in time dimension. Next, we established the FCM model to specify the interrelationship, restrictive correlation and effect degrees between every two elements. During spatial dimension prediction, we introduced thread intelligence data and FCM model to fuse prediction in time and space dimension. At last, we used DARPA2000 data set which in Lincoln Laboratory to verify our method. The result showed that our method can predict the trend of cyber security situation more accurately.

While our method is used in practical application, several problems should be considered. Firstly, how to choose more reasonable indexes to describe cyber security elements in hierarchical index system is needed. Secondly, how to reflect more objective interrelationship, restrictive correlation and effect degrees between every two elements should be

considered. Thirdly, we need to further mining the use of thread intelligence data. The experiment of our method is based on the public DARPA2000 data set, therefore the future research will include improving prediction model, and strengthening the research on other practical application to enhance the versatility of the model.

Acknowledgements

This work is supported in part by the Key Projects in the National Science & Technology Pillar Program of China during the Twelfth Five-year Plan Period under Grant 2015BAK12B03-1, the National High Technology Research and Development Program of China under Grant 2015AA016009 and the National Key R & D Program of China under Grant 2017YFC0803702 and 2017YFB0802302.

References

[1] M. Endsley, Situation Awareness Global Assessment Technique (SAGAT), *The 88th National Aerospace and Electronics Conference*, Dayton, OH, 1988, pp. 789-795.

[2] T. Bass, Intrusion Detection Systems and Multisensor Data Fusion, *Communications of the ACM*, Vol. 43, No. 4, pp. 99-105, April, 2000.

[3] G. Kuang, X. Wang, L. Yin, A Fuzzy Forecast Method for Network Security Situation Based on Markov, *International Conference on Computer Science and Information Processing (CSIP)*, Xi'an, Shaanxi, China, 2012, pp. 785-789.

[4] S. Sun, The Research of the Network Security Situation Prediction Mechanism Based on the Complex Network, *International Conference on Computational Intelligence and Communication Networks (CICN)*, Jabalpur, India, 2015, pp. 1183-1187.

[5] Y. Wang, W. Li, Y. Liu, A Forecast Method for Network Security Situation Based on Fuzzy Markov Chain, in: Y. M. Huang, H. C. Chao, D. J. Deng, J. Park (Eds.), *Advanced Technologies, Embedded and Multimedia for Human-centric Computing*, Springer, 2014, pp. 953-962.

[6] D. Lee, D. Kim, J. Jung, Multi-Stage Intrusion Detection System using Hidden Markov Model Algorithm, *International Conference on Information Science and Security (ICISS)*, Seoul, South Korea, 2008, pp. 72-77.

[7] H. Farhadi, M. AmirHaeri, M. Khansari, Alert Correlation and Prediction using Data Mining and HMM, *The ICS International Journal of Information Security*, Vol. 3, No. 2, pp. 77-101, July, 2011.

[8] M. A. Bode, S. A. Oluwadare, B. K. Alese, A. F. Thompson, Risk Analysis in Cyber Situation Awareness using Bayesian Approach, *International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA)*, London, UK, 2015, pp. 1-11.

[9] G. Yan, R. Lee, A. Kent, D. Wolpert, Towards a Bayesian Network Game Framework for Evaluating DDoS Attacks and Defense, *The ACM Conference on Computer and Communications Security*, Raleigh, North Carolina, 2012, pp. 553-566.

[10] A. A. Ramaki, M. Khosravi-Farmad, A. G. Bafghi, Real Time Alert Correlation and Prediction using Bayesian Networks, *The 12th International Iranian Society of Cryptology Conference on Information Security and Cryptology (ISCISC)*, Rasht, Iran, 2015, pp. 98-103.

[11] F. Castaldo, F. A. N. Palmieri, C. S. Regazzoni, Bayesian Analysis of Behaviors and Interactions for Situation Awareness in Transportation Systems, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 17, No. 2, pp. 313-322, February, 2016.

[12] B. Shi, X. Q. Xie, Research on Network Security Situation Forecast Method Based on D-S Evidence Theory, *Computer Engineering and Design*, Vol. 34, No. 3, pp. 821-825, March, 2013.

[13] C. Wang, L. Fang, D. Wang, Y. Dai, Network Security Situation Awareness System Based on Knowledge Discovery, *Computer Science*, Vol. 39, No. 7, pp. 11-17, July, 2012.

[14] Y. Liu, D. Feng, Y. Lian, K. Chen, D. Wu, Network Situation Prediction Method Based on Spatial-time Dimension Analysis, *Journal of Computer Research and Development*, Vol. 51, No. 8, pp. 1681-1694, August, 2014.

[15] G. Kou, G. Tang, Z. Xu, Research on Revising Conflict Evidence of D-S Evidence Theory in Network Security Situation Awareness, *Computer Science*, Vol. 42, No. 7, pp. 200-203, July, 2015.

[16] Y. Tang, W. Li, J. Yu, X. Yan, Network Security Situational Assessment Method based on Improved D-S Evidence Theory, *Journal of Nanjing University of Science and Technology*, Vol. 39, No. 4, pp. 405-411, August, 2015.

[17] J. Deng, Control Problems of Grey Systems, *Systems & Control Letters*, Vol. 1, No. 5, pp. 288-294, March, 1982.

[18] S. Liu, J. Forrest, Y. Yang, A Brief Introduction to Grey Systems Theory, *IEEE International Conference on Grey Systems and Intelligent Services (GSIS)*, Nanjing, China, 2011, pp. 1-9.

[19] S. Kordnoori, H. Mostafaei, S. Kordnoori, The Application of Fourier Residual Grey Verhulst and Grey Markov Model in Analyzing the Global ICT Development, *Hyperion Economic Journal*, Vol. 2, No. 1, pp. 50-60, March, 2014.

[20] W. Hu, J. Li, X. Chen, X. Jiang, Network Security Situation Prediction Based on Improved Adaptive Grey Verhulst Model, *Journal of Shanghai Jiaotong University (Science)*, Vol. 15, No. 4, pp. 408-413, August, 2010.

[21] L. Chen, Z. Si, R. He, F. Zhou, Network Security Situation Prediction Based on Improved Adaptive Grey Model, *Computer Science*, Vol. 41, No. Z11, pp. 259-262, November, 2014.

[22] C. Tang, Y. Xie, B. Qiang, X. Wang, R. Zhang, Security Situation Prediction Based on Dynamic BP Neural with Covariance, *Procedia Engineering*, Vol. 15, pp. 3313-3317, December, 2011.

[23] C. Tang, X. Wang, R. Zhang, Y. Xie, Modeling and Analysis of Network Security Situation Prediction Based on

Covariance Likelihood Neural, *ICIC'11 Proceedings of the 7th International Conference on Intelligent Computing: Bio-inspired Computing and Applications*, Zhengzhou, China, 2011, pp. 71-78.

[24] R. Zheng, D. Zhang, Q. Wu, M. Zhang, C. Yang, A Strategy of Network Security Situation Autonomic Awareness, *The Second International Conference on Network Computing and Information Security*, Shanghai, China, 2012, pp. 632-639.

[25] Y. Zhang, S. Jin, Predicting Network Security Situation Based on a Combination Model of Multiple Neural Networks, *International Journal of Software and Informatics*, Vol. 8, No. 2, pp. 167-176, August, 2014.

[26] R. B. Myerson, *Game Theory: Analysis of Conflict*, Harvard University Press, 1991.

[27] Q. Wu, S. Shiva, S. Roy, C. Ellis, V. Datla, On Modeling and Simulation of Game Theory-based Defense Mechanisms against DoS and DDoS Attacks, *The Spring Simulation Multiconference*, Orlando, FL, 2010, pp. 1-8.

[28] Y. Shi, R. Li, X. Peng, G. Yue, Network Security Situation Prediction Approach Based on Clonal Selection and SCGM(1,1)c Model, *Journal of Internet Technology*, Vol. 17, No. 3, pp. 421-429, May, 2016.

[29] J. Yang, D. Ryu, J. Baik, Improving Vulnerability Prediction Accuracy with Secure Coding Standard Violation Measures, *The IEEE 2016 International Conference on Big Data and Smart Computing (BigComp)*, Hong Kong, China, 2016, pp. 115-122.

[30] Y. Zhang, X. Yun, M. Hu, Research on Privilege-escalating based Vulnerability Taxonomy with Multidimensional Quantitative Attribute, *Journal of China Institute of Communications*, Vol. 25, No. 7, pp. 107-114, July, 2004.

[31] Z. Baig, K. Salah, Distributed Hierarchical Pattern-Matching for Network Intrusion Detection, *Journal of Internet Technology*, Vol. 17, No. 2, pp. 167-178, March, 2016.

[32] Y. Jia, X. Wang, W. Han, A. Li, W. Cheng, YHSSAS: Large-scale Network Oriented Security Situational Awareness System, *Computer Science*, Vol. 38, No. 2, pp. 4-8, February, 2011.

[33] Z. Gao, Y. Yao, F. Rao, Y. Liu, P. Luo, Predicting Model of Vulnerabilities Based on the Type of Vulnerability Severity, *ACTA Electronica Sinica*, Vol. 41, No. 9, pp. 1784-1787, September, 2013.

[34] H. Peng, Z. Li, Construction of a Categorization Framework for CVE Based on SVM, *Journal of Jishou University (Natural Science Edition)*, Vol. 34, No. 1, pp. 66-71, January, 2013.

[35] O. Alhazmi, Y. Malaiya, I. Ray, Security Vulnerabilities in Software Systems: A Quantitative Perspective, *The 19th Annual IFIP WG 11.3 Working Conference on Data and Applications Security*, Storrs, CT, 2005, pp. 281-294.

[36] K. Chen, D. Feng, P. Su, C. Nie, X. Zhang, Multi-Cycle Vulnerability Discovery Model for Prediction, *Journal of Software*, Vol. 21, No. 9, pp. 2367-2375, September, 2010.

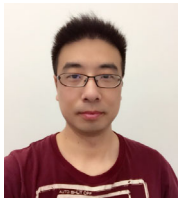
[37] B. Kosko, Fuzzy Cognitive Maps, *International Journal of Man-Machine Studies*, Vol. 24, No. 1, pp. 65-75, January, 1986.

[38] Z. Wang, Research of Network Security Situation Evaluation Based on Index System, Master Thesis, *College of Computer, National University of Defense Technology*, Changsha, China, 2010.

[39] MIT Lincoln Lab, *2000 DARPA Intrusion Detection Scenario Specific DataSets*, http://www.ll.mit.edu/r-d/datasets/2000-darpa-intrusion-detection-scenario-specific-dataset.

[40] Y. Zhang, X. Liu, Weights Learning of Fuzzy Cognitive Maps, *Journal of Chinese Computers Systems*, Vol. 34, No. 5, pp. 1147-1153, May, 2013.

Biographies



Zhijie Fan received his M.S. degree in Zhejiang University, Hangzhou, China, in 2009. He is currently pursuing his Ph.D. degree in the School of Computer Science and Engineering, Tongji University. His research interests include cyber security, cloud computing security, mobile security and communication security.

Zhiping Tan received his B.Sc. degree in Computer Science from UCLA, California, in 2016. He is currently an engineer at Huawei Technologies Co. Ltd. His research interest includes SDN programming for emerging networks such as 5G networks, IoT networks and cloud networks.

Chengxiang Tan received the Ph.D. degree in Engineering from Northwestern Polytechnic University, China in 1994. He is currently the Professor of Computer Science at Tongji University. His research interests include big data, cyber security and privacy preservation, and emerging networks.

Xin Li received the Ph.D. degree in Zhejiang University, Hangzhou, China in 2006. He is currently the Professor of Computer Science at People's Public Security University of China. His research interests include big data, cyber security and video content analysis.