



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Approach to Forecasting Multi-step Attack Based on Fuzzy Hidden Markov Model

¹Zhang yanxue, ²Zhao Dongmei and ³Liu Jinxing

¹College of Mathematics and Information Science,
Hebei Normal University, Shijiazhuang, 050000, China

²College of Information Technology, Hebei Normal University, Shijiazhuang, 050000, China

³The First Aeronautics College of PLAAF, Xinyang, 464000, China

Abstract: We integrate the approach of forecasting multi-step attack, the association rule, fuzzy evaluation and Hidden Markov Model (HMM) and support the method of forecasting multi-step attack based on fuzzy Hidden Markov Model. Firstly, we fuse raw alerts into super alert. Then we obtain the initial state matrix by the probability of the attack, determine the state transition matrix by the association rule and obtain the observation matrix by fuzzy evaluation. Finally, we recognize the alert belonging to attack scenarios with the Forward algorithm of HMM and forecast the next possible attack sequence with the Viterbi algorithm of HMM. Simulation experiments results verify the validity of the approach.

Key words: Multi-step attack, alert processing, association rule, fuzzy evaluation, hidden markov model

INTRODUCTION

Currently, the network security situation is increasingly sophisticated and the multi-step network attack has become the mainstream of network attack. 2012 Chinese Internet network security reports released by the National Computer network Emergency Response technical Team Coordination Center of China (CNCERT/CC) shows that worms, Distributed Denial of Service (DDoS) (Xie *et al.*, 2013) and other multi-step network attacks account for 60% of overall network attacks. Multi-step attack (Yuan, 2010) means the attackers apply multiple attack steps to attack the security holes of target itself and achieve the devastating blow to the target (Wang *et al.*, 2007). In the multi-step attack, there is a causal relationship between multiple attack steps and also have the characteristics of the property of time sequence (Chen and Yan, 2011) and the uncertainty on steps (Zhai and Zhou, 2011) and so on.

Current research on the approaches to forecasting multi-step attack behaviors, mainly includes four types: (1) The approach to forecasting multi-step attack based on the antecedents and consequences of the attack (Wang and Cheng, 2005). It applies the precursor subsequent relationship of the event, to forecasting the attacker wants to implement attacks in the near future. Because of the complexity and the diversity of the attack behaviors, this approach is difficult to achieve. (2) The approach to forecasting multi-step attack based on Hierarchical Colored Petri Nets (HCPN) (Wu *et al.*, 2008; Yan *et al.*, 2006), it applies the raw alerts by Petri Nets and

considers the attack intention is inferred by raw alerts. But this approach focus on the detection of multi-step attack behaviors (Zhai and Zhou, 2011). (3) The approach to forecasting multi-step attack based on Bayes game theory (Cao *et al.*, 2007a, b). It could forecast the probability that the attackers choose to attack and the probability that the defenders choose to defend in the next stage rationally. However, in current study, only two person game model is established, so this approach has some limitations. (4) The approach to forecasting multi-step attack based on attack intention (Chen and Yan, 2011; Zhang, 2007; Wang and Cheng, 2005), It uses extended-directed graph to describe the logical relationship between attack behaviors and forecasts the next stage by the logical relationship. The shortcoming of this approach is that it is difficult to determine the matching degree (threshold) of the multi-step attack.

In order to achieve the effective forecast of multi-step attack, In this study we propose a new approach to forecasting multi-step attack-the approach to forecasting multi-step attack based on fuzzy Hidden Markov Model. Firstly, we fuse raw alerts into super alert. Then we obtain the initial state matrix by the probability of the attack, determine the state transition matrix by the association rule and obtain the observation matrix by fuzzy evaluation. Finally, we recognize the alert belonging to attack scenarios with the Forward algorithm of HMM and forecast the next possible attack sequence with the Viterbi algorithm of HMM. Simulation experiments results verify the validity of the approach.

MODEL OF FORECASTING MULTI-STEP ATTACK BASED ON FUZZY HIDDEN MARKOV MODEL

Hidden Markov Model (HMM) (Faeiz *et al.*, 2010; Lee *et al.*, 2008) usually used to deal with the problems related to the time sequence and it has been widely used in speech recognition, signal processing, bioinformatics and other fields. Recent years, Hidden Markov Model is also used in the field of intrusion detection field.

Hidden Markov Model is characterized by the following:

- S: The number of states in the model
- V: The number of observation symbols per state
- A: The state transition matrix
- B: The probability distribution of V
- p: The initial state probability distribution of this model

Based on the characteristics of Hidden Markov Model and the concealment, difficult to observe and forecast of multi-step network attack behaviors, so we propose a fuzzy Hidden Markov Model and realize the recognition and forecasting of multi-step attack.

The process of the approach is listed as follows: Firstly, we fuse raw alerts into hyper alert. Then we obtain the initial state matrix by the probability of the attack, determine the state transition matrix by the association rule and obtain the observation matrix by fuzzy evaluation. Finally, we recognize the alert belonging to attack scenarios with the Forward algorithm of HMM and we forecast the next possible attack sequence with the Viterbi algorithm of HMM. Simulation experiments results verify the validity of the approach. The flow chart is shown in Fig. 1.

The model of recognizing and forecasting multi-step attack based on fuzzy Hidden Markov Model is shown in Fig. 2.

RELATED WORK

Raw alerts processing: Based on the characteristics of the raw alerts associated with semantic analysis, the format of the raw alert is defined as: RawAlert(RawAlert_ID, RawAlert_Type, Source_IP, Destination_IP, Start/End_Time), the format of the hyper alert is defined as: HyperAlert(HyperAlert_ID, HyperAlert_Type, Source_IP, Destination_IP, Start/End_Time, Alerts_Count).

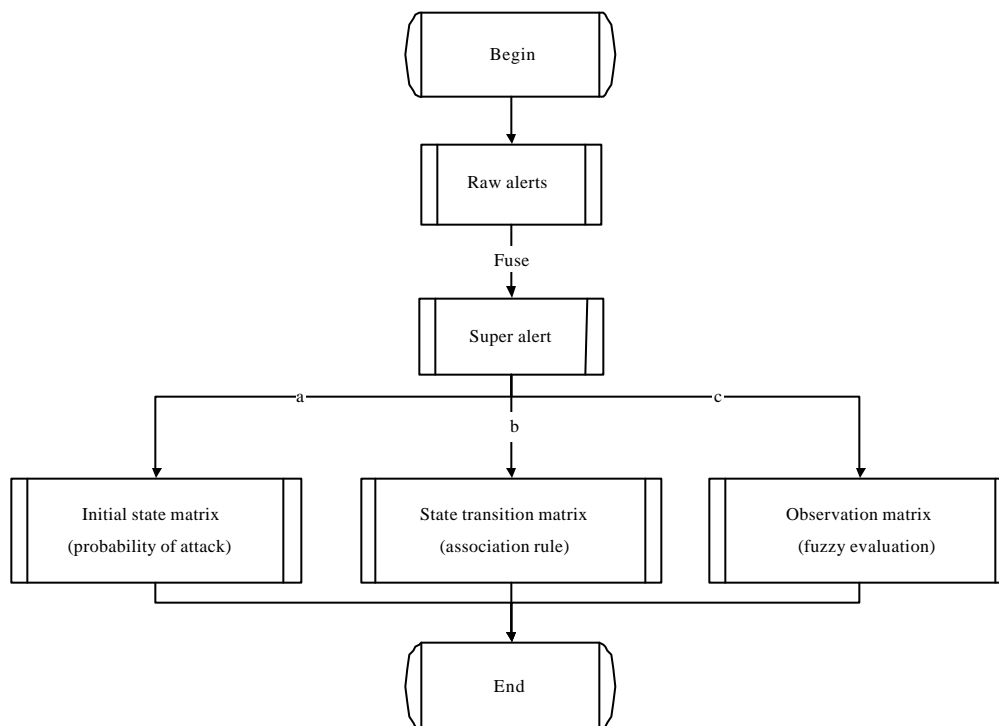


Fig. 1: Flow chart of forecasting multi-step attack based on fuzzy hidden markov model

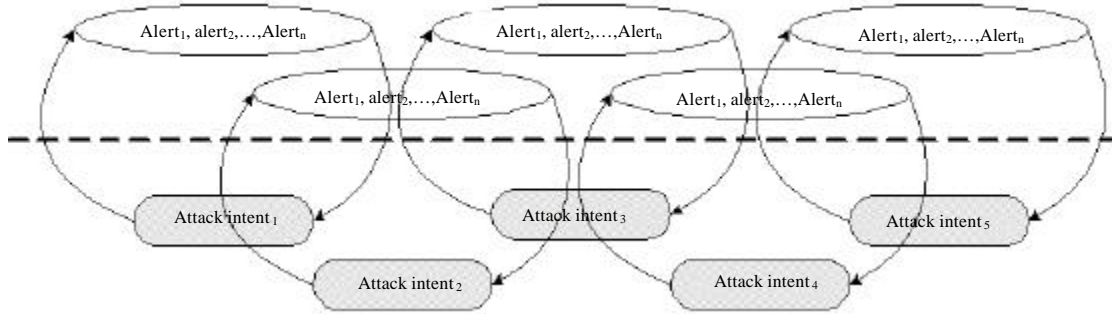


Fig. 2: Model of recognizing and forecasting multi-step attack based on fuzzy hidden markov model

Table 1: Raw alerts

Alert ID	Alert name	Source IP	Destination IP	Start/end time
1	ICMP Echo Reply	172.16.112.1	172.16.113.168	05:18/05:18
2	ICMP Echo Reply	172.16.112.2	172.16.113.168	05:18/05:18
...
M	ICMP Echo Reply	172.16.112.m	172.16.113.168	05:31/05:31

Table 2: Hyper alert

HyperAlert ID	Hyperalert name	Source IP	Destination IP	Start/end time	Alert Count
001	ICMP Echo Reply	172.16.112.*	172.16.113.168	05:18/05:31	m

Raw alerts fuse into hyper alert, rules are defined as follows:

- In the raw alerts, if the values of the attributes-RawAlert_ID and Start/End_Time are different and the values of the other three attributes are the same, we will recognize these raw alerts are repeating alerts of an event and leave only one of them and discard the rest of them
- We fuse raw alerts that the values of these two attributes-Source_IP and Destination_IP are the same into hyper alert (Zhai and Zhou, 2011). As is shown in Table 1 and 2

Determination of state transition matrix-association rules: By the recent research of association rules (Zhang *et al.*, 2008; Zhang, 2007), we find that association rules and mining technique can help us calculating the state transition matrix in multi-step network attack behaviors which effectively reflects the dependencies between attack intentions.

We assume (1) There are two attack intentions in the multi-step attack X, intention i, intention j (2) a_{ij} are the state transition probability between two attack intentions and $a_{ij} = p(\text{intention } i \rightarrow \text{intention } j)$. If the intention items has association rule: Intention i \rightarrow intention j, we say there is a state transition relationship between intention i and intention j. In this study, a_{ij} is calculated as follows:

$$a_{ij} = p(\text{intention } i \rightarrow \text{intention } j) = \frac{\text{count}(\text{intention } i \rightarrow \text{intention } j)}{\text{count}(\text{intention } i)} \dots \dots *$$

In (*), count (intention i \rightarrow intention j) represents the number of intention i and intention j are simultaneously in an intention items and intention j followed intention i. count(intention i) represents the number of intention i occurred in the intention set. For instance, as is shown in Fig. 3 and according to (*), the calculation results are shown in Table 3.

Determination of observation-fuzzy evaluation: Fuzzy evaluation is a method based on fuzzy mathematics comprehensive evaluation. This method depends on the membership degree of fuzzy mathematics theory, transform qualitative into quantitative evaluation. That is to say, using fuzzy mathematics to make a comprehensive decision to a thing which is constrained by a number of factors. Fuzzy evaluation can effectively deal with its native subjectivity during the evaluation process, as well as deal with the encountered fuzzy phenomenon objectively (Zhao *et al.*, 2009).

We assume (1) B: The relative weight of each alert under a certain criterion. (2) A: weight set. (3) R: The membership degree matrix. Where:

$$B = A \circ R = (a_1, a_2, \dots, a_m) \circ \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & r_{mn} \end{pmatrix}$$

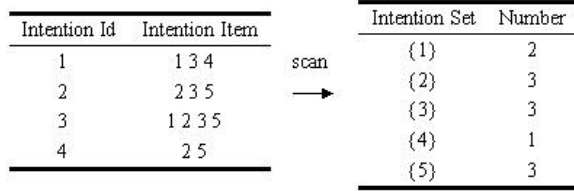


Fig. 3: Initialization

Table 3: Results

	{1}	{2}	{3}	{4}	{5}
{1}	0	1/2	1/2	0	0
{2}	0	0	2/3	0	1/3
{3}	0	0	0	1/3	2/3
{4}	0	0	0	0	0
{5}	0	0	0	0	0

Table 4: Definition of impact level (Zhao, 2007)

Impact Level	Description
v ₁	Negligible
v ₂	Small
v ₃	General
v ₄	Serious
v ₅	Key

◦ is the fuzzy operator. In this study, we will use fuzzy operator M(•, ⊕). Example below.

- Factor set: U = {A₁, A₂, A₃, A₄, A₅, A₆, A₇, A₈, A₉, A₁₀, A₁₁, A₁₂, A₁₃}. A is short for Alert. This rule also applies to the others parts of this study
- Evaluation set: V = {v₁, v₂, v₃, v₄, v₅}. The definition of v_i is shown in Table 4

According to Table 4, experts make a probabilistic evaluation to the factor set U. Each expert decide its impact on the various factors, the impact probability is one of v₁, v₂, v₃, v₄, v₅. Combine with the evaluation on each expert, we calculate for each alert impacts the completion of attack intention and get the membership matrix R as the observation matrix:

$$R = \begin{bmatrix} 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.490 & 0.490 & 0.020 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.200 & 0.200 & 0.200 & 0.200 & 0.200 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.660 & 0.170 & 0.170 \end{bmatrix}$$

Setting the weights of v₁, v₂, v₃, v₄, v₅, the weight order is 1/25, 3/25, 5/25, 7/25, 9/25. According to the equation B = A◦R, we get the relative weights: [0.040, 0.059, 0.059, 0.024, 0.040, 0.040, 0.040, 0.040, 0.040, 0.280, 0.238, 0.010, 0.010]. After the Normalization, we get the vector [0.043, 0.043, 0.065, 0.043, 0.065, 0.043, 0.043, 0.043, 0.043, 0.043, 0.043, 0.304, 0.259, 0.011, 0.011].

From the above results, we can obtain that A₁₀ and A₁₁ have the biggest impact to the completion of the attack intention. When we detect the same alert in this type of multi-step attack in the future, we should increase our efforts to prevent and avoid the occurrence of such attacks.

Algorithm to recognizing and forecasting multi-step attack based on hidden markov model-forward algorithm and viterbi algorithm: The steps of Forward Algorithm are as follows:

Step 1: Initialization:

$$\alpha_1(i) = \pi_i b_i(o_1), \text{ where } 1 \leq i \leq N \tag{1}$$

Step 2: Iterative calculation:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(o_{t+1}) \tag{2}$$

Where:

$$(1 \leq t \leq T-1, 1 \leq j \leq N)$$

Step 3: Termination condition:

$$p(O|\lambda) = \sum_{i=1}^N \alpha_T(i) \tag{3}$$

Among them, λ is the given HMM model, O is the observation sequence and O = {o₁, o₂, ..., o_k}.

The steps of Viterbi Algorithm are as follows:

Step 1: Initialization:

$$\delta_1(i) = \pi_i b_i(o_1), \text{ where } 1 \leq i \leq N \tag{4}$$

$$\psi_1(i) = 0 \tag{5}$$

Step 2: Iterative calculation:

$$\delta_t(j) = \max(\delta_{t-1}(i) a_{ij}) b_j(o_t) \tag{6}$$

$$\begin{aligned} &\text{where } 1 \leq i \leq N \\ &\psi_t(j) = \arg \max(\delta_{t-1}(i) a_{ij}) \end{aligned} \tag{7}$$

Step 3: Termination conditions:

$$P^* = \max(\delta_T(i)) \tag{8}$$

$$q_T^* = \arg \max (\delta_T(i)) \tag{9}$$

Step 4: The optimal path:

$$q_t^* = \Psi_{t+1}(q_{t+1}^*) \tag{10}$$

Where:

$$t = T-1, T-2, \dots, 1$$

SIMULATION EXPERIMENT AND ANALYSIS

The data set is used in the simulation experiment is an attack scenario testing data sets LLDOS1.0 (inside) provided by DARPA (Defense Advanced Research Projects Agency) in 2000. We extract two kinds of multi-step attack from it, they are DDoS multi-step attack and FTP Bounce multi-step attack (Zhao *et al.*, 2009). According to section 2 and we establish two fuzzy Hidden Markov Model, they are DDoS_HMM and FTP Bounce_HMM. The parameters of fuzzy Hidden Markov Model are shown as follows.

- Alerts and attack intentions of DDoS_HMM are shown as follows:
- IPSweep-A1: ICMP Echo Reply
- SadmindPing-A2: RPC portmap sadmind request UDP, A3: RPC portmap Solaris port query udp request, A4: RPC sadmind UDP Ping
- SadmindExploit-A₅: RPC portmap Solaris sadmind port query udp request, A₆: RPC port sadmind request UDP, A₇: RPC sadmind UDP, A₈: RPC sadmind UDP Netngt_Proc_Service Client_Domain Overflow attempt, A₉: RPC PORTMAP Solaris sadmind port query udp portmapper sadmind port query attempt
- InstallDDoSTools-A₁₀: Rservices rsh root
- Launch DDoSAttack-A₁₁: SNMP AgentX/tcp request, A₁₂: SNMP trap tftp, A₁₃: SNMP request tcp

The initial state matrix p, state transition matrix A and observation matrix B of DdoS_HMM is shown from Table 5-7.

- Alerts and Attack Intentions of FTP Bounce_HMM are shown as follows:
 - IPSweep-A₁: ICMP Echo Reply, A₂: ICMP Ping NMAP
 - PortScan-A₃: Scan NMAP TCPM, A₄: Scan synscan port scan

Table 5: Initial state matrix p

Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
0.250	0.750	0.000	0.000	0.000

Table 6: State transition matrix A

	State1	State2	State3	State4	State5
State1	0.000	1.000	0.000	0.000	0.000
State2	0.000	0.177	0.823	0.000	0.000
State3	0.000	0.228	0.688	0.028	0.056
State4	0.000	0.000	0.000	0.750	0.250
State5	0.000	0.000	0.000	0.000	0.000

Table 7: Observation matrix B

	Alert1	Alert2	Alert3	Alert4	Alert5	Alert6	Alert7	Alert8	Alert9	Alert10	Alert11	Alert12	Alert13
State1	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
State2	0.000	0.490	0.490	0.020	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
State3	0.000	0.000	0.000	0.000	0.200	0.200	0.200	0.200	0.200	0.000	0.000	0.000	0.000
State4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
State5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.660	0.170	0.170

Table 8: Initial state matrix p

Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
0.667	0.333	0.000	0.000	0.000

Table 9: State transition matrix

	State1	State2	State3	State4	State5
State1	0.600	0.400	0.000	0.000	0.000
State2	0.000	0.823	0.177	0.000	0.000
State3	0.000	0.000	0.625	0.375	0.000
State4	0.000	0.000	0.000	0.750	0.250
State5	0.000	0.000	0.000	0.000	1.000

Table 10: Observation matrix B

	Alert1	Alert2	Alert3	Alert4	Alert5	Alert6	Alert7	Alert8	Alert9	Alert10
State1	0.250	0.750	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
State2	0.000	0.000	0.118	0.882	0.000	0.000	0.000	0.000	0.000	0.000
State3	0.000	0.000	0.000	0.000	0.625	0.375	0.000	0.000	0.000	0.000
State4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
State5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.833	0.167

- FTPExploit-A₅: FTP anonymous login attempt, A₆: FTP anonymous ftp login attempt
- A₇: FTP forward
- RhostModify-A₈: FTP rhosts
- LaunchFTPBounceAttack-A₉: Rservices rsh root, A₁₀: Rservices rlogin root

The initial state matrix p, state transition matrix A and observation matrix B of FTP Bounce_HMM is shown from Table 8-10.

Example 1: When we received alerts "ICMP Echo Reply" and "RPC portmap Solaris sadmind port query udp

request”, according to the forward algorithm of Hidden Markov Model, we will obtain the probability based on DDoS_HMM and FTP Bounce_HMM, respectively:

$$P(\text{Alerts} | \text{DDoS_HMM}) = 0.1225$$

$$P(\text{Alerts} | \text{FTP Bounce_HMM}) = 0.0079$$

We can see from the above results, $P(\text{Alerts} | \text{DDoS_HMM}) > P(\text{Alerts} | \text{FTP Bounce_HMM})$, that is to say, the ongoing multi-step attack behavior is likely to be DDoS attack.

Example 2: When the console receives the alert sequence {Alert₁, Alert₂, Alert₃, Alert₄}, we can obtain the completed attack sequence (Faeiz *et al.*, 2010; Lee *et al.*, 2008; Wu *et al.*, 2008). That is to say, now completed attack is the previous three attack intentions-IPSweep, SadminPing and SadminExploit, next attack intention would be InstallDDoSTools.

CONCLUSION

By the current research on the approaches to forecasting multi-step attack behaviors, we integrate the approach of forecasting multi-step attack, the association rule, fuzzy evaluation and Hidden Markov Model (HMM) and support the method of forecasting multi-step attack based on fuzzy Hidden Markov Model. By this approach we can recognize and forecast the multi-step attack better. Simulation experiments results verify the validity of the approach.

ACKNOWLEDGMENTS

The authors would like to thank the reviewers for their detailed reviews and constructive comments which have helped improve the quality of this study. This work was supported by The National Natural Science Foundation of China No. 60573036, Hebei Science Fund under Grant No F2013205193 and Hebei science supported planning projects No. 12213514D.

REFERENCE

Cao, H., Q.Q. Wang, Z.Y. Ma and P. Luo, 2007. Attack prediction model based on dynamic bayesian games. *Comput. Appl.*, 27: 1545-1547.
Cao, H., Q.Q. Wang, Z.Y. Ma and P. Luo, 2007. Attack prediction model based on static Bayesian game. *Appl. Res. Comput.*, 24: 122-124.

Chen, C. and B.P. Yan, 2011. Network attack forecast algorithm for multi-step attack. *Comput. Eng.*, 5: 172-174.
Faeiz, A., A. Monis, U.A. Irfan, J.C. Anrea and M. Parvin, 2010. MARS: Multi-stage attack recognition system. *Proceedings of the 24th International Conference on Advanced Information Networking and Applications*, April 20-23, 2010, Perth, Australia, pp: 753-759.
Lee, D.H., D.Y. Kim and J.I. Jung, 2008. Multi-stage intrusion detection system using hidden markov model algorithm. *Proceedings of the International Conference on Information Science and Security*, January 10-12, 2008, Seoul, Korea, pp: 72-77.
Wang, L., 2007. Study on method of network multi-stage attack plan recognition. Master's Thesis, Huazhong University of Science and Technology, Hubei.
Wang, Z.L. and X.P. Cheng, 2005. An Attack predictive algorithm based on the correlation of intrusions alerts in intrusion response. *Comput. Sci.*, 32: 144-146.
Wu, R.Y., W.G. Li and H. Huang, 2008. An attack modeling based on hierarchical colored petri nets. *Proceedings of the International Conference on Computer and Electrical Engineering*, December 20-22, 2008, Phuket, pp: 918-921.
Xie, B.L., S.G. Jiang and Q.S. Zhang, 2013. Application-layer DDoS attack detection based on request keywords. *Comput. Sci.*, 40: 121-125.
Yan, F., H. Huang and X.C. Yin, 2006. A detection algorithm for multi-step attack based on CTPN. *Chinese J. Comput.*, 29: 1383-1391.
Yuan, C., 2010. Research on multi-step attack detection method based on GCT. Master's Thesis, Jilin University, Jilin.
Zhai, G.Q. and S.Y. Zhou, 2011. Construction and implementation of multistep attacks alert correlation model. *J. Comput. Applic.*, 31: 1276-1279.
Zhang, S.H., 2007. Research on network security early warning technology based on hidden markov model. Master's Thesis, PLA Information Engineering University, Henan.
Zhang, S.H., Y.D. Wang and J.H. Han, 2008. Approach to forecasting multi-step attack based on HMM. *Comput. Eng.*, 34: 131-133.
Zhao, D.M., 2007. Study on the risk assessment quantitative method of information security. Master's Thesis, Xidian University, Shanxi.
Zhao, D.M., J.X. Liu and J.F. Ma, 2009. Risk assessment of information security using fuzzy wavelet neural network. *J. Huazhong Univ. Sci. Technol.*, 37: 43-45.