Network Security Risk Assessment Method based on the Improved Hidden Markov Model

Xinlei Li*

Henan Normal University; Henan Xinxiang 453007,China

***Abstract* - In this paper, we concentrate on the problem of assessing the network security risk via an improved Hidden Markov model. As is well known that Hidden Markov model is belonged to one of the regime-switching models, which have been successfully exploited to model time series data. However, the standard Hidden Markov model cannot effectively to represent the structure of holding times for a specific hidden state. To solve this problem, we design an improved Hidden Markov model, in which parameters are calculated based on parameter space's gradient. Furthermore, considering the security state of a host cannot be observe directly, the security state of a given host is estimated by analyzing the network security event sequences. Next, we define four possible security states of a host, that is, "Good", "Probed", "Attacked", and "Compromised", and possible state transitions are illustrated as well. Afterwards, the risk of the network security can be computed via summing risks of all hosts. To testify the effectiveness of the proposed algorithm, we design four experimental schemes based on five types of hosts, including: Web server, Email server, FTP server, Database server and Personal computer. Experimental results demonstrate that compared with BAG based method, the performance of our proposed algorithm is much closer to the true value.**

Keywords* - *Network security risk, Hidden Markov model, Hidden state, Transition matrix, Cost vector

I. INTRODUCTION

With the fast development of information technology and computer network, computer networks have played an important role in the aspects of politics, economy, and so on. However, the network security problems are very important for us to solve[1]. To solve the increasingly difficult network security problems, a lot of network security defense and control method are proposed. Considering the technology of network security assessment may effectively find potential vulnerabilities and provide the security condition of network system, it has become one of the most important methods to prevent dangers. As one of most important security defense approaches, network security risk assessment technology is utilized to evaluate security risks before the dangerous event happening[2][3]. Particularly, the suitable risk management methods are chosen in terms of the risk assessment results. Hence, effective and efficient network security risk assessment approaches are very important for network or information system security construction. Based on researching on the related works, in this paper, we proposed an improved hidden Markov model to tackle the network security risk assessment problem.

Furthermore, the individual vulnerabilities are not important for network security, however, the effective integrated vulnerabilities may greatly destroy network security[4]. In recent years, there are several potential security problems when sharing resources in Internet, and network security problems have received more and more attentions when launching network research[5]. However, most of the existing methods of network security risk assessment ignore the relationships between vulnerabilities.

The major contributions of this paper lie in that we introduce the Hidden Markov model in the network security risk assessment problem. The hidden Markov model refers to a statistical Markov model in which the system being modeled can be defined as a Markov process with several hidden states. Particularly, the hidden Markov model can be represented as the simplest dynamic Bayesian network[6-8].

The paper is organized as follows. In the next section, we introduce the related works about hidden Markov model and give analysis. Section 3 illustrates an overview of the Hidden Markov model. Section 4 proposes a network security risk assessment approach based on the improved Hidden Markov model. Afterwards, experiments are conducted in section 5, and this paper is concluded in section 6.

II. RELATED WORKS AND ANALYSIS

Hidden Markov model is belonged to one of the regime-switching model, which has been a dominant method to the modeling of the type of time series data. In this section, we will analyze how the hidden Markov model widely used in many application fields, and then discuss why we should improve the standard hidden Markov model.

Manandhar et al. modified the standard HMM model via a multiple-instance learning, which utilized an unordered set of HMM sequences at a specific alarm location. In this paper, the set of sequences is regarded as positive only when there is a sequence in a target sequence. Meanwhile, with the proposed multiple-instance learning model, the bags and the corresponding labels are exploited to train the target and non-target HMM model at the same time[9].

Charles et al. applied hidden Markov models to infer vessel activities in the snow crab. In this paper, the hidden Markov model classified three behavioral states in the VMS data. The proposed method testifies that behavioral variables can contribute to the standardization of catch similar to classical trip and vessel variables[10].

Travers demonstrates that a finite hidden Markov model with path-mergeable states the block estimates of the entropy rate converge exponentially rapidly. Particularly, the authors describe that the path-mergeability attribute is asymptotically typical in the space of hidden Markov topologies and simply testable as well[11].

Wang et al. presented a trajectory-based outlier discovering method via model training and model-based likelihood computation through a multi-dimensional Hidden Markov Model. Furthermore, the authors explore the possibility and feasibility of exploiting the given method to real-time outlier discovering[12].

Gassiat et al. supposed that finite state space stationary hidden Markov models in the situation where the number of hidden states is uncertain. Furthermore, the authors proposed a frequentist asymptotic evaluation of Bayesian analyzing approaches. Based on the conditions on the prior, the authors also defined a consistent Bayesian estimator of the number of hidden states[13].

Lin et al. designed a new sequential Monte Carlo algorithm based on the idea that non-linear filtering of continuous-time jump Markov processes. The main innovations of this paper lie in that particle learning is used to minimize particle degeneracy and then utilize the analytical jump Markov framework[14].

Apart from the above works, hidden Markov model has also used in other fields, such as Blind Categorical Deconvolution[15], Longitudinal data analyzing[16], Computers and Electronics in Agriculture[17], Signal Modeling[18], speaker identification in noise environment[19], Degradation modeling and monitoring[20].

III. OVERVIEW OF THE HIDDEN MARKOV MODEL

As is well know that Hidden Markov model is flexible models and it is very useful in different range of applications. However, there is a serious limitation in Hidden Markov model, which refers to inability of HMM to explicitly structure the holding times of a specific hidden state.

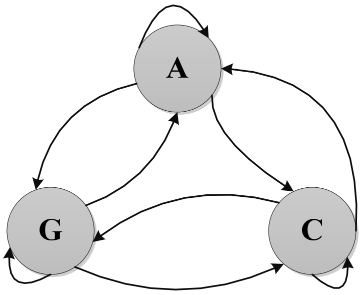

Fig. 1 A description of a fully connected Markov model

As is shown in Fig.1, we provide a fully connected Markov model with three different states, including G (good), A (under attack), and C (compromised), that is $S = \{G, A, C\}$.

Supposing that $\{Z_t\}_{t=1}^{\infty}$ denotes a Markov chain, in which Z_t refers to hidden state the time point t. Furthermore, there are N possible values $S = \{S_1, S_2, \cdots, S_N\}$. Moreover, the transition matrix contains the transition probabilities among the hidden states, and it is defined by the following equation.

$$\alpha_{ij} = P\left(Z_{t+1} = S_j \middle| Z_t = S_i\right) \quad (1)$$

$$\text{s.t. } i, j \in \{1, 2, \cdots, N\} \text{ and } \sum_{j=1}^{N} \alpha_{ij} = 1 \quad (2)$$

On the other hand, $\{Y_t\}_{t=1}^{\infty}$ refers to the observation symbol when the time point t is reaching, and the values of observation sequences are represented as $V = \{V_1, V_2, \cdots, V_M\}$. The emission matrix $B = \{b_j(V_k)\}$ can represent the probability distribution for each state as follows.

$$b_j(V_k) = P\left(Y_t = V_k \middle| Z_t = S_j\right) \quad (3)$$

where the number of $\sum_{j=1}^{N} b_j(V_k)$ is equal to one. To represent the hidden process, we propose the definition of initial state probability as follow.

$$\pi = \{\pi_1, \pi_2, \cdots, \pi_N\} \quad (4)$$

where π_i is equal to $P(Z_1 = S_i)$ and $\sum_{i=1}^{N} \pi_i$ is equal to one. Assuming that a given set of space evolution $Q = \{q_1, q_2, \cdots, q_T\}$ and a set of possible symbols

$V=\{v_1,v_2,\cdots,v_T\}$, the sequence probability can be assessed by the following equation:

$$\begin{aligned} P(V|M) &= \sum_Q P(V,Q) = \sum_Q P(Q)P\langle V|Q\rangle \\ &= \sum_Q \prod_{t=1}^{T} P(q_t|q_{t-1}) \cdot \prod_{t=1}^{T} p(v_t|q_t) \end{aligned} \tag{5}$$

IV. NETWORK SECURITY RISK ASSESSMENT BASED ON THE IMPROVED HIDDEN MARKOV MODEL

As is well known that Hidden Markov model is widely used in many different applications by maximizing the marginal probability of the observations based on the hidden states. Hence, it is a generative approach and need not fully exploit the inter-class discriminative information from training set. A suitable method to tackle this problem is to put it into a discriminative learning system. Therefore, in this section, we proposed an improved Hidden Markov model, in which the parameter of the HMM is computed by parameter space's gradient.

In our proposed improved Hidden Markov model, gradient of the logarithm of the probability which is corresponding to parameter λ is computed as follows.

$$U(x/\alpha) = \nabla\alpha_{i,j} \cdot \log P\left(\frac{X}{\lambda}\right) = \frac{\xi(x,s_j)}{\alpha_{i,j}} - \xi(s_j) \tag{6}$$

$$U(x/\beta) = \nabla\beta_{i,j} \cdot \log P\left(\frac{X}{\lambda}\right) = \frac{\xi(s_i,s_j)}{\beta_{i,j}} - \xi(s_j) \tag{7}$$

where the symbol $\xi(x,s_j)$ means the number of times at which a specific symbol x is obtained by a state s_j in a particular sequence, and $\xi(s_i,s_j)$ refers to the frequency of the joint occurrence between two different states s_i and s_j based on a same sequence. On the other hand, α and β refer to the forward and backward variables respectively, and $\xi(s_i)$ is the frequency of the state s_i which is happening in a sequence.

Supposing that U_X is the kernel vector of the improved Hidden Markov model, and the similarity of two sequences with the improved Hidden Markov model can be computed as follows.

$$K(X,Y) = K(U_X,U_Y) \tag{8}$$

where the function $K(\square)$ denotes any kind of standard kernels for the support vector machine.

The main innovations of this paper lie in that we creatively introduce the Hidden Markov model to the problem of network security risk assessment. In this problem, we cannot observe the security state of a host, however, we can estimate the security state via analyzing the network security event sequences which are corresponding to the given host. We define security states of the host as $S=\{G,P,A,C\}$. Particularly, four states G,P,A,C represent 1) "Good": in good state, and has not been attacked, 2) "Probed": has been probed by attackers, 3) "Attacked": the given host has been attacked, and the invasion probability is increasing, 4) "Compromised": the host has been attacked and be in the most dangerous state. The possible state transitions are illustrated in Fig.2 as follows.

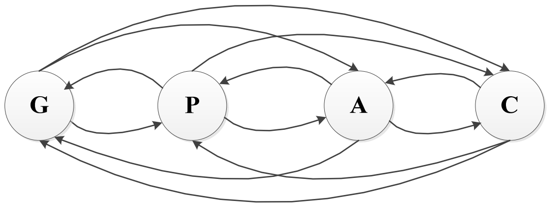

Fig. 2 Illustration of possible host security state transitions

Next, we suppose that a host can observe M types attacks (denoted as $V=\{v_1,v_2,\cdots,v_M\}$), and the attacking sequence is represented as $O=\{o_1,o_2,\cdots o_t,\cdots,o_T\}$, and o_t is belonged to set V. To compute the security state, the proposed hidden Markov model contains a triple $\lambda=(Trans,Obs,\pi)$, where $Trans$ denotes the state transition matrix, and $Trans_{ij}$ means the probability when the host state is S_i is at time t and state become S_j at time $t+1$. This process can be represented as follows.

$$\begin{aligned} &Trans_{ij} = P(qt+1=sj|qt=si) \\ &s.t.\quad i,j\in[1,N] \end{aligned} \tag{9}$$

Obs denotes the observing matrix, and Obs_{nm} means the probability of host being observed in the state S_n at time t is equal to a_m.

$$Obs_{nm} = P\left(Ot = am \middle| qt = sn\right)$$
$$s.t.\ \ n \in \left[1, M\right] and\ m \in \left[1, N\right] \tag{10}$$

The initial state π is a vector, which can represent the probability of the state for each host ($\pi = \left(r_1, r_2, \cdots, r_N\right)$). The state distribution at time t is represented as $r_t = \left\{r_t\left(i\right)\right\}, i \in \left[1, N\right]$, and the equation of state distribution probability is defined as follows.

$$r_t\left(i\right) = P\left(q_t = s_i \middle| y_t\right) \tag{11}$$

Afterwards, we introduce a cost vector θ to represent the risk value for each state of a given host, and then quantitative analysis can be conducted, and in the following experiment, $\theta = \{1, 30, 60, 100\}$. Based on the above definition, the current risk value of the given host is defined as follows.

$$R = \sum_{i=1}^{N} r_i \cdot c_i \tag{12}$$

Then, the risk of the whole network can be computed by summing the risk of each host in the network.

V. EXPERIMENTS

In this section, we design an experiment to make performance evaluation. Particularly, services provided by servers and hosts are listed in Table.1.

TABLE.1 SERVICES PROVIDED BY SERVERS AND HOSTS

Type	Function	Service
Web server	Web page service	Internet Information Services (IIS)
Email server	Email service	SMTPD
FTP server	Providing file service for Web server	FTPD, SSHD
Database server 1	Providing database service for Web server	Oracle
Database server 2	Providing database service for Email server	Oracle
Computer 1	Logging in the FTP server	Starting the function of remote desktop
Computer 2	Logging in the FTP server	None

As is shown in Table.1, five types of devices are included, which are 1) Web server, 2) Email server, 3) FTP server, 4) Database server and 5) Personal computer. Afterwards, the relationships between server and host are illustrated in Table.2 as follow:

TABLE.2 RELATIONSHIPS BETWEEN SERVERS AND HOSTS

Source	Destination	Assess	Port
Arbitrarily	Web server	HTTP	80
Arbitrarily	Email server	SMTP	25
Web server	FTP server	FTP	21
Web server	Database server 1	SQL	1433
Email server	Database server 2	SQL	1433
Other computer	Computer 1	RDP	1435
Computer 1	FTP server	SSH	22
Computer 2	FTP server	SSH	22

To test the effectiveness of our proposed network security risk assessment algorithm, we design four schemes to make performance evaluation as follows.

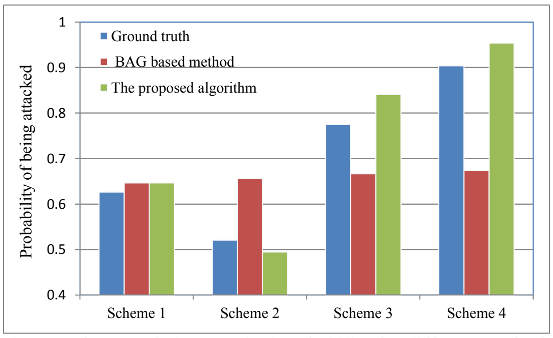

Fig. 3 Web server being attacked probability for different methods

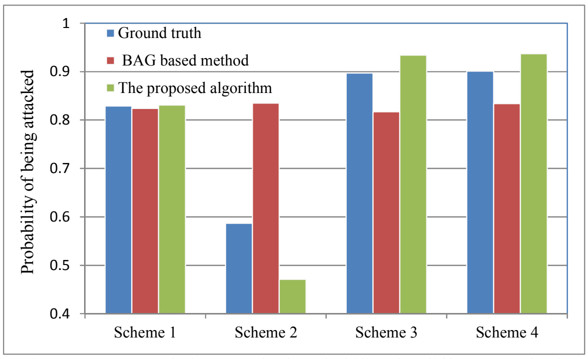

Fig. 4 FTP server being attacked probability for different methods

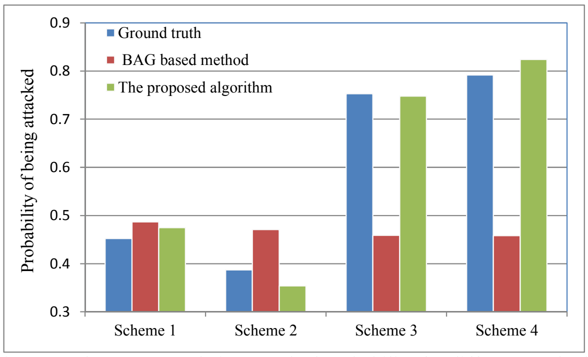

Fig. 5 Database server 1 being attacked probability for different methods

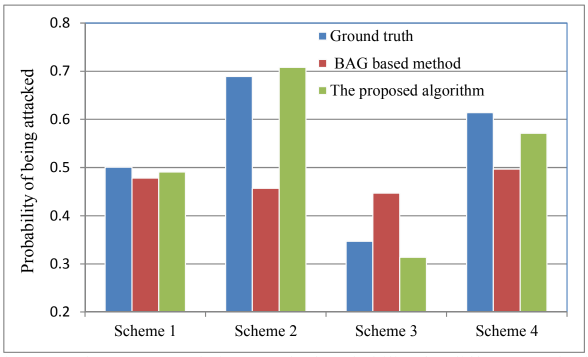

Fig. 6 Database server 2 being attacked probability for different methods

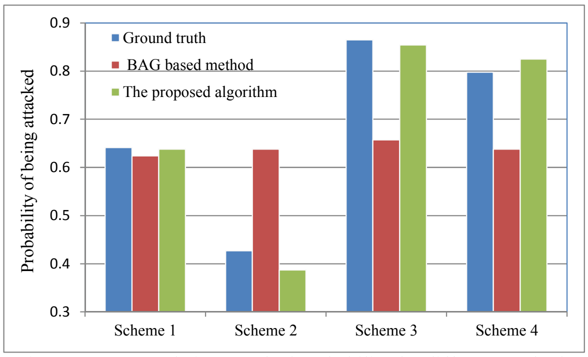

Fig. 7 Computer 1 being attacked probability for different methods

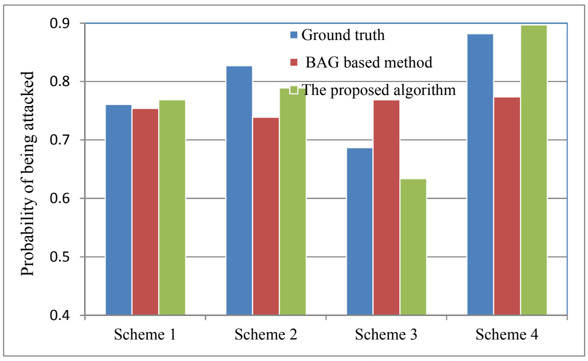

Fig. 8 Email server being attacked probability for different methods

Scheme 1: Web server, FTP server, Database server 1,2, Computer 1,2, and Email server have almost the same asset value and importance. Hence, attacker can obtain almost the same attacking profit as well, and each host has the same probability of being attacked.

Scheme 2: Web server only publishes normal pages. When Web server goes into the fault state, it has a little effect on the whole network system. Furthermore, the data and files used in Web server are memorized in FTP server and Database server 1. There are no important information which are memorized in Computer 1 and 2.

Scheme 3: In this network, Web server undertakes the main task of the network. When failure happens in the Web server, the network may be influenced greatly. On the other hand, important and files which the Web server require are memorized in FTP server and Database server 1.

Scheme 4: These basic settings of this scheme are just similar to scheme 3, and the difference between them lies in that Email server and Database server have been suffered from frequent attacks.

To make performance comparison, the method in paper [21] (denoted as BAG based method) is used to compare with our proposed algorithm. Moreover, we collect a dataset by collect the attacking behaviors in a long period time, and make the attacking probability as the ground truth. Experimental results for different kinds of host are shown in Fig.3 to Fig.8.

As is shown in Fig.3 and Fig.8, we can see that attacking probability of our proposed algorithm is much closer to the ground truth, and the detailed information is illustrated in Table.3.

TABLE.3 ERROR RATE OF THE NETWORK SECURITY RISK ASSESSMENT

	Web server	FTP server	Database server 1	Database server 2	Computer 1	Email server	Average
BAG based method	17.2%	14.8%	27.7%	21.5%	13.5%	8.9%	17.27%
The proposed algorithm	5.6%	7%	4.6%	5.3%	6.1%	3.8%	5.4%

Table.3 demonstrates that the error rates of network security risk assessment for BAG based method and the proposed algorithm are 17.27% and 5.4% respectively. Therefore, we can know that the proposed can effectively assess the network security risk.

VI. CONCLUSIONS

This paper studies on the problem of assess the network security risk based on an improved Hidden Markov model. Particularly, we propose an improved Hidden Markov model, in which parameters are obtained based on parameter space's gradient, and the security state of a given host is estimated by analyzing the network security event sequences. Afterwards, four security states are defined as "Good", "Probed", "Attacked", and "Compromised". Then, the risk of the network security is obtained by summing risks of all hosts. Experimental results verify the effectiveness of the proposed algorithm.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGMENT

This work is supported by the Program for Science and Technology Innovation Talents in Universities of Henan Province (No. 13HASTIT040).

REFERENCES

[1] Marsadek M., Mohamed A., "Risk based security assessment of power system using generalized regression neural network with feature extraction", Journal of Central South University, 2013, 20(2): 466-479.

[2] Viduto Valentina, Maple Carsten, Huang Wei, Lopez-Perez David, "A novel risk assessment and optimisation model for a multi-objective network security countermeasure selection problem", Decision Support Systems, 2012, 53(3): 599-610.

[3] AF Sanchez M. M., "Security risk assessments in public transport networks", Proceedings of The Institution of Mechanical Engineers Part F-Journal of Rail and Rapid Transit, 2011, 225(F4): 417-423.

[4] AF Ralston P. A. S., Graham J. H., Hieb J. L., "Cyber security risk assessment for SCADA and DCS networks", ISA TRANSACTIONS, 2007, 46(4): 583-594.

[5] Yang Fu-Hong, Chi Chi-Hung, Liu Lin, BE Yang LT, Jin H, Ma J, Ungerer T, "A risk assessment model for enterprise network security", Autonomic and Trusted Computing, 2006, 4158: 293-301.

[6] Economou Theodoros, Bailey Trevor C., Kapelan Zoran, "MCMC implementation for Bayesian hidden semi-Markov models with illustrative applications", Statistics and Computing, 2014, 24(5): 739-752.

[7] Arakawa Toshiya, Tanaveh Akira, Ikeuchi Shiho, et al. "A male-specific QTL for social interaction behavior in mice mapped with automated pattern detection by a hidden Markov model incorporated into newly developed freeware", Journal of Neuroscience Methods, 2014, 234: 127-134.

[8] Skewes-Cox Peter, Sharpton Thomas J., Pollard Katherine S., DeRisi Joseph L., "Profile Hidden Markov Models for the Detection of Viruses within Metagenomic Sequence Data", PLOS ONE, 2014, 9(8), Article No.e105067.

[9] Manandhar Achut, Torrione Peter A., Collins Leslie M., Morton, Kenneth D., "Multiple-Instance Hidden Markov Model for GPR-Based Landmine Detection", IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(4): 1737-1745.

[10] Charles Colin, Gillis Darren, Wade Elmer, "Using hidden Markov models to infer vessel activities in the snow crab (Chionoecetes opilio) fixed gear fishery and their application to catch standardization", Canadian Journal of Fisheries and Aquatic Sciences, 2014, 71(12): 1817-1829.

[11] Travers Nicholas F., "Exponential bounds for convergence of entropy rate approximations in hidden Markov models satisfying a path-mergeability condition", Stochastic Processes and Their Applications, 2014, 124(12): 4149-4170.

[12] Wang Chen, Lin Hongzhi, Jiang Hongbo, "Trajectory-based multi-dimensional outlier detection in wireless sensor networks using Hidden Markov Models", Wireless Networks, 2014, 20(8): 2409-2418.

[13] Gassiat Elisabeth, Rousseau Judith, "About the posterior distribution in hidden Markov models with unknown number of states", BERNOULLI, 2014, 20(4): 2039-2075.

[14] Lin Junjing, Ludkovski Michael, "Sequential Bayesian inference in hidden Markov stochastic kinetic models with application to detection and response to seasonal epidemics", Statistics and Computing, 2014, 24(6): 1047-1062.

[15] Lindberg David Volent, Omre Henning, "Blind Categorical Deconvolution in Two-Level Hidden Markov Models", IEEE Transactions on Geoscience and Remote Sensing, 2014, 52(11): 7435-7447.

[16] Lagona Francesco, Jdanov Dmitri, Shkolnikova Maria, "Latent time-varying factors in longitudinal analysis: a linear mixed hidden Markov model for heart rates", Statistics in Medicine, 2014, 33(23): 4116-4134.

[17] Aparna U., Pedersen Lene Juul, Jorgensen Erik, "Hidden phase-type Markov model for the prediction of onset of farrowing for loose-housed sows", Computers and Electronics in Agriculture, 2014, 108: 135-147.

[18] Huda Shamsul, Yearwood John, Togneri Roberto, "Hybrid Metaheuristic Approaches to the Expectation Maximization for Estimation of the Hidden Markov Model for Signal Modeling", IEEE Transactions on Cybernetics, 2014, 44(10): 1962-1977.

[19] Shahin Ismail, "Novel third-order hidden Markov models for speaker identification in shouted talking environments", Engineering Applications of Artificial Intelligence, 2014, 35: 316-323.

[20] Cholette Michael E., Djurdjanovic Dragan, "Degradation modeling and monitoring of machines using operation-specific hidden Markov models", IIE TRANSACTIONS, 2014, 46(10): 1107-1123.

[21] Poolsappasit Nayot, Dewri Rinku, Ray Indrajit, "Dynamic Security Risk Management Using Bayesian Attack Graphs", IEEE Transactions on Dependable and Secure Computing, 9(1): 61-74, 2012.