Intrusive Behavior Analysis Based on Dynamic Honeynet and Multidimensional Hidden Markov Model

Chang-Lung Tsai[1*] and Min-Hsiung Hung[1,2]

[1] *Department of Computer Science, Chinese Culture University, Taiwan*
[2] *Department of Electrical and Electronic Engineering, CCIT, National Defense University, Taiwan*

ABSTRACT

In this paper, an intrusion behavior analysis mechanism based on timely updated dynamic honeynet and multidimensional hidden Markov model has been proposed. In the proposed scheme, three modules named as monitor module, track module, and analysis module are developed. All of the intrusive behavior and their corresponding intrusion trails are traced and all of the intrusion information are collected and analyzed through the above modules combined with the deposit of different pheromones for indexing corresponding record of intrusion and recording the whole process of documents being accessed to verify the attacked target and behavior mode. Among the developed honeynet, the honey pots, databases, directories, and parameters of system and security application are automatically updated in timely to reach the goal of attracting the hackers. Experimental results demonstrate that the proposed mechanism possesses good efficiency and performance in tracing networking intrusion and behavior analysis.

Keywords: information security, intrusion detection system, hidden Markov model, honeynet

以動態誘捕網路與多維度馬可夫隱藏模型為基之入侵行為分析

蔡昌隆[1*]　洪敏雄[1, 2]

[1]中國文化大學資訊科學學系
[2]國防大學理工學院電機電子工程學系

摘　　要

本論文提出基於即時更新之動態誘捕網路與多維度之隱藏馬可夫鏈以追蹤駭客入侵之路徑與行為，其架構包含監控、追蹤與分析等三個模組。所有駭客入侵之路徑與行為均透過此模型搭配螞蟻演算法配予相應費洛蒙權重以追蹤記錄駭客入侵路徑及竊取資料庫文件之過程，以驗證其攻擊目標與行為模式。本文設計之誘捕網路，其蜜罐、資料庫、目錄與系統相關安全參數等均採自動化動態即時更新以達誘捕目的，實驗證明本機制確具良好分析與追蹤效能。

關鍵詞：資訊安全，入侵偵測系統，馬可夫隱藏模型，誘捕網路

文稿收件日期 99.7.2;文稿修正後接受日期 100.1.5;*通訊作者
Manuscript received July 2, 2010; revised January 5, 2011; * Corresponding author

Ⅰ. INTRODUCTION

As the application of cloud computing, Internet and wireless networking emerge, the information security events increase significantly. Although almost all of the networking system of the enterprises and institutes are mounted with firewall, antivirus, antispyware, intrusion detection and prevention system, or with further special setting of parameters to protect the networks and information systems from diversity intrusion and network attacking. However, there still thousands of network attacking and intelligent criminal happened everyday such as unauthorized login, access to confidential file or classified database, privilege escalation, malware, and etc. Networking attacks and information security events happened day after day in everywhere. Not only millions of general documents are sniffed every day, but also those confidential parameters, classified files and with authorized database are under the severe threaten of access and cracking. Unfortunately, currently no one can predict when the intruders will stop network attacking.

As for gaming, most of the player might take direct attacks instead of passive defense. However, network attacking is still illegal up to dates in the real world. Therefore, protect networking security through active attacking is impossible.

To solve the significant information security problem, a lot of researchers are focused on the development of the intrusion detection system (IDS) and intrusion prevention system (IPS) [1] [2] [3] [4] [5]. Currently, most of the IDSs detect the network intrusion based on anomaly statistical analysis, pattern recognition, signature matching, protocol identification and application. Although all of the signatures, parameters, confidential values, and audit data are provided for intrusion analysis, currently the IDSs still possess higher false positive and false negative probability [6] [7]. In order to raise networking security, well understanding the intrusive activity and analyzed their intrusion techniques and the corresponding intrusive rationale through an optimal designed monitoring mechanism is one of the most adopted approaches. In which, thoroughly and effectively understanding the thought and behavior of intrusion is the foundation to construct more strong and concrete IDS. It is also the key to provide optimal information protection and achieve the goal of security assurance. Moreover, recently some of the IDSs have combined with honey pots or honey nets to raise their performance. As to the design of honey pots, how to raise the mutual interaction between intruders and systems is the key factor of successfully to attract those intruders to visit the honey pots and preventing or delay to attack the critical information systems.

In [8], Pfleeger et al. focus the networking security from the viewpoint of insiders. They proposed a mechanism to describe the insiders and their corresponding actions based on the consideration of organization, environment, system, and individuals. Although some of the real examples of insider action such as hard drive removal, stolen intellectual property, tax fraud, and proliferation of e-mail responses are provided, the viewpoint is focused on insiders. However, except those unwelcomed access from insiders, a lot of networking attack are initiated from the Internet.

In [9], Joshua pointed out that NIDS used to log suspicious activities are potentially signified the security violations. In addition, there is shortcoming of wider acceptability in computer security to succinctly process audit logs based on intrusion information integrated analysis and high level of expertise to validate each alert are needed frequently. In order to effective detect the intrusion, well understanding the behavior of intrusion is a must. In [10], Renuka et al. proposed a hybrid framework for behavioral prediction of network attacking by utilization of honeypot and dynamic rule creation. Their output is a dynamic list of attacks which are generated by a classifier and queued in the honeypot for behavior analysis and pattern matching. Although the proposed rationale again enhanced the significance of intrusive behavior analysis, blacklisting is not an optimal approaching. The reason is that there are thousands of attacking patterns occurred in the Internet every day. As the patterns of the black list increase, the recognition and matching loading increase. Relatively, the performance of intrusion detection and protection might decrease and the information protection cost becomes increased.

In this paper, a novel intrusive analysis model is proposed and developed based on the design of honey-net. The reason is that there exists value-added benefit of deploying the honey pots and honey nets for intrusive behavior analysis. That is the reason that a dynamic updating honey net is developed in our proposed mechanism and it is performed with the combination of intrusion detection system to exactly capture the access trail and behavior of the intrusion. The traversal of intrusion is then tracked and analyzed by utilized the ant colony algorithm with the deposit of different pheromones. In addition to the deposit of the pheromone, there is additional corresponding weight designed to indicate the importance of each distinct nodes, files, directories, and databases inside the operated networking system. More pheromones mean higher degree of attraction and more attempts occurred. That also means higher probability of information being accessed or stolen. The intrusion is manipulated according to the deposit and evaporation of pheromone and the corresponding weight. Since distinct cracker possesses different capability of network attacking, in order to exactly, correctly, and completely understand and categorize the capability of the intruders, the weighting scheme is dynamically tuned to fit the real capability of intruders. After then, the enhanced defense system can be constructed based on the analysis of the intrusive behavior.

The rest of the paper is organized as follows. In Section 2, the proposed intrusive behavior analysis mechanism is presented. The concept of ant colony algorithm for analyzing the trail of intruders is addressed in the section. The rationale of multidimensional hidden Markov model for integrated intrusive information analyzing is also explicated in this section. Experimental results are illustrated in Section 3 to demonstrate the feasibility and validity of our proposed intrusive behavior analysis mechanism. Finally, concluding remarks are given in Section 4.

Ⅱ. THE PROPOSED INTRUSIVE BEHAVIOR ANALYSIS SCHEME

Although there are a lot of IDSs developed, the techniques for network attacking are innovated very fast. Therefore, the requirement and satisfaction of the performance of the IDSs for all of the information security administrators is always endless. Especially, the performance of finding out the kernel cause, fixing the leakage and vulnerabilities, and solving the information security events in real time is the most important consideration. To enhance the efficiency of IDSs and reduce the information loss and operation risk from information security threat, completely comprehension the necessity and their behavior of intruders is a must. Currently, the weighting of parameters defined in most of the IDSs does not come along with dynamic timing factor and real time updating. In addition, the sniffed and audit information is not collected completely. Thus, as the IDS suffering a new intrusive event, the evaluation of risk, influence, significance of targets, and the emergency response might be the same as before. Therefore, to remedy the information security problem and reduce the higher false positive and false negative probability occurred in the IDSs become very important.

As the technology emerging, the hacking and intrusive technique also upgrades. Considering the intrusive activity might occur through many different stages combined with different strategy and versatile techniques, Lee et al have proposed multi-stage IDS based on hidden Markov model to remedy this problem [11]. In their proposed scheme, the intrusion is detected based on those agents deployed on each stage. Nevertheless, most of the attacks are initiated through distributed channels or performed based on organized model. Although numbers of detecting sensors are mounted and controlled by each agent, the intrusive data collected or analyzed from a single detection agent might not be integrated completely for association analysis. Therefore, the developed IDS must possess the capability of monitoring a huge networking system and extract the intrusive evidence from the integrated information that collected from distributed agents.

2.1 The Proposed Mechanism

In this section, an intrusive behavior based analysis scheme is proposed in the paper. In which, three modules named as monitor module, track module, and analysis module are developed as shown in Figure 1.

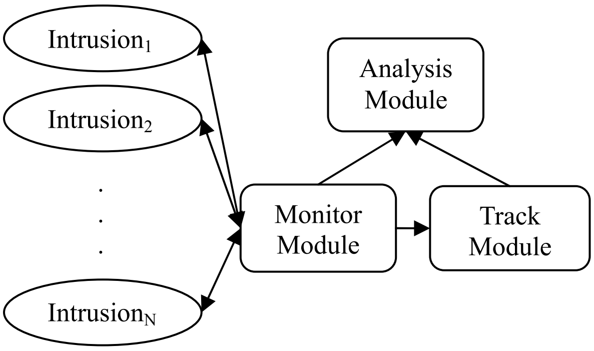


Fig. 1 The diagram of the proposed intrusive behavior analysis mechanism.

The intrusive behavior and hacking activity is detected by an intrusion detection module and analyzed based on Ant colony optimization algorithm [12] [13].

All of the behavior information including the susceptive intrusion, misuse, and normal access are collected in the monitor module. In addition, the system log, security parameter, protocol setting, service, audit information, and etc. will also be recorded.

After then, the complete path and traverse of each distinct access including legal access and illegal intrusion will be deposit with different pheromones and weights in the track module. Finally, all of the data will be transferred to the analysis module for further integrated analysis of intrusive behavior based on ant colony algorithm.

2.2 Rationale of Ant Colony Algorithm for Intrusive Analysis

Colony is one of the nature behaviors of ant. They inform each other of the discovered trail by depositing some pheromone. The trail that more ants walk through there more pheromones deposit. However, the pheromone will gradually evaporate according to the time past. There are a lot of researches which are focus on the application of the ant colony algorithm including solving those dynamical and stochastic problems in the field of network management. The behavior of most of the network intrusions are also possessed the characteristic of dynamical and stochastic initially. However, an organized network intrusion or distributed network attacking could be diagnosed through statistical analysis.

Practically, an ant will wander randomly in the initial. After finding food, the ant will return back to their colony and deposit pheromone on the passed trails. Therefore, other ants can easily identify the path and follow the trail to instead of wandering randomly. After then, the pheromone that has been deposited on the trail will start to evaporate and the attraction is then decreased. The traverse of a new intrusion is just similar as the behavior an ant during finding food. If the intruders found something interest after the discovering, most of them will visit again. Otherwise, they will seldom or almost not to visit those noninterest targets again. That is the reason that the ant colony algorithm is very suitable for applying to solve the problem of intrusive behavior analysis.

The iterative manipulation of ant colony algorithm for intrusive behavior analysis is described as the following steps.

1. Initial Phase: The behavior of most of the individual hackers is same as the ant activity in this phase.

(1) Suppose an intrusion is occurred randomly around the whole networking system.

(2) In the beginning, the pheromone is set to be zero.

(3) As the intruders start to discover, the deposit of pheromone for successfully attempt is indicated as either PH_{hc} or PH_c. If the attempt is failed, there still has the deposit of pheromone to record the attempt which will be indicated as either F_PH_{hc} or F_PH_c.

(4) In general, if the discovered area is belonging to general environment without any confidential setting, the attempt will be indicated as PH_g.

(5) The entire possible deposit amount of the pheromones is tabulated in Table 1 according to the significance of the discovered areas.

2. Discovering phase: If there exists something interest files, documents or things, and etc. in the network for intruders, most of the situations they will discover the whole network similar as the ant finds food source.

In the discovering phase, an intruder might take some kind of actions such as the following:

(1) Construct communication tunnel.

(2) Setup the backdoors.

(3) Advanced search: Keep on surfing the websites or searching the database.

(4) Covert: Hide the trail and omit the record.

Leave immediately: Similar as the ants go back to their nest. All of the above possible action will leave pheromone on the trail more or less based on the proposed mechanism as tabulated in Table 1.

Table 1. Illustrating of different pheromone parameter to indicate different significance of visit area

Pheromone Item		Index	Signifi-cance	Notation
Attempt Success	PH_{hc}	1	Highly confidential or classified	indicates the intruders successfully visit or access into the highest confidential/classified area
	PH_c	2	Confidential or classified	indicates the intruders successfully visit or access into confidential/classified area
Aattempt Failure	F_PH_{hc}	3	Highly confidential or classified	indicates the intruders fail to visit or access into the highest confidential/classified area
	F_PH_c	4	Confidential or classified	indicates the intruders fail to visit or access into confidential/classified area
Attempt	PH_g	5	General/without confi-dential	indicates the visit of intrusion into those area without the request of permission or authourity

3. Revisit phase: Under normal situation, as the ants find the food source, the other ants will be attracted to follow on this trail. The density of the pheromones will then rise until all of the food has been carried out and the pheromones evaporate gradually.

No matter the intruders are belonging to individuals or some kind of organized groups, once the intruders found something interest, they will visit or access again with highly probability. Organized intrusions are just like the colony visiting.

In order to record the movement of the intruders in the whole network system, each device, servers, files, directories, databases, and etc. are treated as an individual node and mounted with a caption to indicate their distinct properties. A simplified example is listed in Table 2. In addition, the number indexing is utilized to tell the different devices that belonged to the same set such as router 1 and router 2 will be indicated as N_{r1} and N_{r2}.

Table 2. Node indexing for each distinct device, servers, files, directories, databases, and etc.

Hardware		Software	
Index	Notation	Index	Notation
N_s	server	N_{os}	operating system
N_r	router	N_{ap}	application program
N_{sw}	switch	N_{db}	database
N_{fw}	firewall	N_f	Files document
N_t	terminal	N_{sp}	security parameter

Since there are a lot of devices and nodes within the whole networking system, thus the movement of intruders from i^{th} node to j^{th} node is a stochastic process. To describe the movement, probability model is applied and the movement of the intruders is measured by utilizing the following equation.

$$\Pr_{i,j} = \frac{\left|PH_i - PH_j\right|^{\alpha}}{\sum_{i,j}\left|PH_i - PH_j\right|^{\alpha}} \qquad (1)$$

where PH_i indicates the pheromone amount of the i^{th} node and PH_j indicates the pheromone amount of the j^{th} node and α is a power parameter used for tuning. In the above equation, how many pheromones should be deposited is defined according to Table 1. The variation of the density of the pheromone on each node, PH_DN_i, can be described as the following equation.

$$PH\_DN_i = \sum_{i,\tau} PH_i \bullet \alpha^{\tau} \qquad (2)$$

where i is the *index* which denotes the i^{th} kind of pheromone as tabulated in Table I, α is an evaporation factor which ranges from zero to one, i.e. $0<\alpha\leqq 1$, and τ is a timing factor which denotes the duration between two consecutive attempts of the i^{th} node.

The pheromones of each trail can be summarized as the following equation.

$$PH\_DN_{i,j} = \sum_{j} \sum_{i,\tau} PH_i \bullet \alpha^{\tau} \qquad (3)$$

where j denotes those distinct nodes on the trail.

The pheromone is updated according to the variation of the attempts, evaporating factor and time duration.

2.3 The Architecture of Honey-net

A honey pot is performed as a decoy on servers or systems which possess a lot of vulnerabilities to attract the intelligentsia or attackers. The goal of the deployed honey pot is set for gathering intrusive information in order to apparently trace the intrusion. Once the honey pot has been invaded, intrusive alert will be generated.

Traditionally, the honey pot is categorized into low interaction and high interaction. The design of honey pot is also divided into physical honey pot and virtual honey pot. Most of the honey pots are deployed at each possible networking nodes or devices such as in the zone between Internet and DMZ, or DMZ to Intranet, or just deployed inside the Intranet of an institute or enterprise as shown in Figure 2.

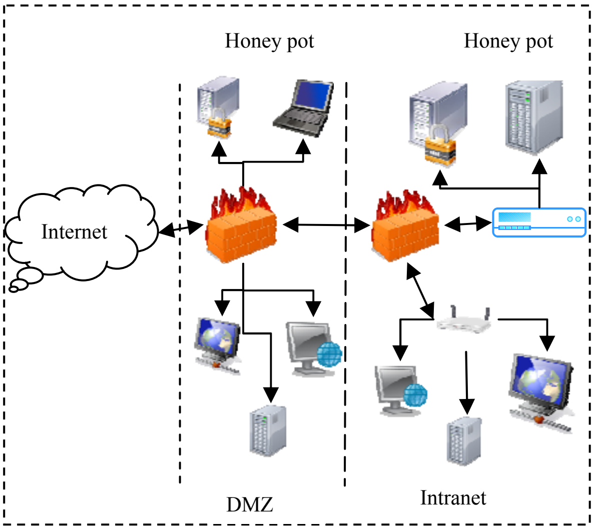


Fig. 2 Illustration of the design of a honeynet.

To implement a honeynet, some of the basis parameters and items for developing a honey pot are listed in the following.

- The architecture of operating environment.
- The kernel operating systems.
- Those opened and closed communication ports.
- The decoyed application programs.
- General documents, sensitive files, confidential files, or encrypted files, and etc.
- The definition of privileged or shared files, directories, databases, and etc.
- The distribution of servers including web server, mail server, database server, and etc. and all of the possible terminals and end user systems.
- Special developed trail for attracting and guiding the intruders enter into the trap.

After the development of a honey pot, all of the distinct items will be given an initial weight to indicate its significance. As the system is initiated, all of the attempts such as discovering the associative ports, files, directories, paths, and databases or even trying to cracking the password and accessing those unauthorized or privileged documents will be deposited with different amount of pheromone as mentioned in previous subsection. All of the completely journey of each different kind of intruder and its corresponding information including the possible IP address, user ID, time stamps, and traverse will be recorded with different amount of pheromone. Therefore, more attempts mean more pheromone will be deposited. After then, the density of the accumulated pheromone of each node evaporates. As the time pass longer, the density of the deposited pheromone of any distinct node will be lesser. All of the collected data will be manipulated and analyzed by utilizing the ant colony algorithm to trace the trail of intrusion and analyze the behavior of the network intruders.

In order to effectively analyze the intrusion especially for those distributed and organized attacking, those targets that have been discovered with higher attempts will be emphasized. In addition, since most of the intruders will not discover the target with a same trail, the cross-correlation measurement between each distinct trail will be performed for integrated diagnosis. Therefore, even though those trails are different the final target is still the same.

Although some of the honey pots are developed, currently the performance of interaction and the frequency of updating the system parameters are still not satisfied.

Those shortcomings induce the result that most of the intruders are not easily attracted to revisit. Therefore, in order to achieve the goal of well control the intrusion, some significant requirement for developing the proposed honeypot is listed as the following:

(1) Each node within the whole networking system is dispatched with different weighting to indicate their different and distinguished significance. In addition, the weighting is closely related with their cracking index as listed in Table 3.

(2) The content of each of the discovered nodes should be updated immediately after they are discovered in order to raise the attraction for intruders.

(3) The security and system parameters of each of the discovered nodes should be dynamically enhanced and their security class

should be raised a little higher right after they are discovered to attract the intruders to access and challenge again.

(4) As the security setting raised and parameter changed, the corresponding weighting should also be updated synchronously.

Different difficulties of access the privileged or classified database is designed to attract the intruders and challenge them in order to measure their attack capabilities. Moreover, different roles possess different authorized privilege to access those rows or columns that as part of classified table that stored in the databases.

Table 3. Illustrating of distinct significance and weighting of nodes within the networking system

Cracking	Weighting Range	Possible Application	Notation
Higest Difficulty	0.8~1.0	● Administrator ● Authority ● IDS parameters	Highly confidential or classified
Difficulty	0.5~0.8	● Database of access lists ● Privileged access area ● Important information	Confidential or classified with random picture verification
Easy	0.2~0.5	● Application program with lower importance ● Opening access with basic authentication	With simple ID and password verification
No Secure	0~0.1	● General files or directories ● Open source ● Open access area	Without any confidential setting

2.4 Multidimensional Hidden Markov Model

All of the digital systems can be described as a set of N different discrete sequences such as {A_{s1}, A_{s2},…, A_{sN}} according to the preset sampling time unit/period. It is also can be mapped into another model and recognized as state transition.

As regarding the intrusion, the intrusive path and discovered nodes or files also can be mapping into discrete sequences. Considering the tracing of intrusive path, the transversal of an intrusion can be treated as state transition and represented in the form of {A_{si}} where "I" is a timing factor and is defined as 0≦i≦N. After then, all of the collected information sent from each detection agent and the deployed sensors can be recorded and demonstrated in discrete model. Finally, the track of an intrusion, i.e. the state transition of each agent, can be considered as a set of discrete sequences described at time factor.

Regarding the intrusive activity, the transversal extracted from each intrusion can be formed similar as a change of state under regular sampling frequency. In order to obtain the benefit of exactly analysis, each state of the intrusive activity can be corresponded with a probability of transversal direction which also forms a probability set associated with each state.

As to the set of states and probabilities, the utilized time instants are associated with the changes of each state and will be denoted as a time sequence as t =1, 2,… and the actual state at time t will be denoted as I_t.

The completely probabilistic description of the intrusion requires the specification of current state (at time t) and all of the predecessor states. Since the steps of an intrusion is very close related, under such condition, the description of the HMM can be simplified to just keep the current and the predecessor state as shown in the following equation.

$$P\lfloor i_t = As_j \mid i_{t-1} = As_i, i_{t-2} = As_k, \ldots \rfloor = P\lfloor i_t = As_j \mid i_{t-1} = As_i \rfloor \quad (1)$$

The set of state transition probabilities of an intrusion is defined as the following equation:

$$p_{i,j} = P\lfloor i_t = As_j \mid i_{t-1} = As_i \rfloor, 1 \le i, j \le N \quad (2)$$

$$p_{i,j} \ge 0$$

$$\sum_{j=1}^{N} p_{ij} = 1 \quad (3)$$

From the above equation, the output results of the process are a set of states corresponding to time instants. Each of the states is also associated with an observable physical information security event. Therefore, the above stochastic process can be categorized as an observable Markov model [3] [10]. However, not all of the information security events are belong to observable Markov model, thus the utilization of higher order or multidimensional HMM for intrusion analysis would be more appropriately in fact. Therefore, each detection agent in the proposed mechanism will be assigned with a respectively HHM to descript its status and the detecting results of all of the sensors dispatched under the control of this agent will thoroughly be collected together. After then, all of the HHM of multi-stage will be integrated.

In order to correctly and exactly detect those associated or distributed attacking, autocorrelation and cross-correlation manipulation will be applied for further analysis.

(1) Autocorrelation measurement: The detecting results belonging to each detection agent will be provided for autocorrelation measurement. The input data for manipulation is the entire information collected from the dispatched sensors under the control of this detection agent. The measurement of autocorrelation for each independent detection agent is computed as the following equation.

$$A_r\left(S_i, S_j\right) = \frac{E\left[\left(S_i - m_{S_i}\right)\left(S_j - m_{S_j}\right)\right]}{\sigma_{S_i}\sigma_{S_j}} \qquad (4)$$

where S_i represents the i^{th} sensor and m_{si} is the mean and σ_{si} is the variance of i^{th} sensor.

(2) Cross-correlation measurement: The detecting results collected from each detection agent will be provided for cross-correlation manipulation. The input data for manipulation is the entire information collected from the dispatched sensors under the control of each detection agent. The measurement of cross-correlation between different detection agents are computed as the following equation.

$$C_r\left(A_u\left(S_i\right), A_v\left(S_j\right)\right) = \frac{E\left[\left(A_u\left(S_i\right) - m_{A_u\left(S_i\right)}\right)\left(A_v\left(S_j\right) - m_{sA_v\left(S_j\right)}\right)\right]}{\sigma_{A_u\left(S_i\right)}\sigma_{A_v\left(S_j\right)}} \qquad (5)$$

where $A_u(S_i)$ represents the i^{th} sensor of the u^{th} detection agent and $m_{Au(Si)}$ is the mean and $\sigma_{Au(Si)}$ is the variance of i^{th} sensor of the u^{th} detection agent.

To response exactly, the proposed IDS mechanism is designed to be reconfigured under operating condition. The data process diagram is shown as Figure 3.

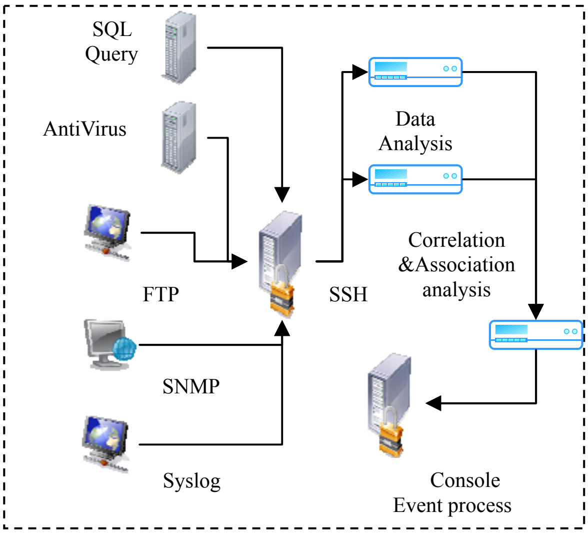


Fig. 3 The diagram of security event process.

In which, all of the information such as Syslog, SNMP (Simple Network Management Protocol) data, SQL query, record of antivirus and firewall and etc. will be collected and transferred through SSH mode to DA (data analysis) unit and finally passed to the Correlation and Association unit for advanced and integrated analysis. All of the analyzed results will presented to the Console unit for taking corresponding event process.

As a new virus or attacking occurred, the new signature will also be added to the database for updating the Expert module of the IDS in real time to keep the system under optimal operating condition. Moreover, in order to tell the different from those individual network intruders and the organized network intruders, spatial domain and frequency domain analysis has been adopted in this paper.

Ⅲ. EXPERIMENTAL RESULTS

In [14], Ahmad et al. evaluated the performance of IDS by utilizing analytic hierarchy process on some approaches such as statistical, rule based, expert system, pattern recognition and artificial neural network (ANN). In their work, artificial neural network has been concluded as more suitable approach for IDS adoption according to those factors such as regular updating, detection rate, false positive, false negative, suitability and flexibility. However, ANN might be suitable for post processing than real time intrusion detection under normal condition. In [15], Li et al. has adopted fuzzy logic with HMM and they gave new definition for the parameters. To evaluate their performance, 1998 DARPA data set has been used and compared with Vemuri and Pradeep. However, the data sets are not available currently.

In order to implement the simulation of Internet hacking, a honeynet is implemented as shown in Figure 4. In which, some hosts that mounted with Snort IDS within the local area network are assigned to play the role of target to form a generalized IDS for experiment. One server is acted as management. In addition, inside each hosts there are sensors and application programs which is implemented to collected the intrusive activities under secret hidden mode. Therefore, the hackers will not be conscious of the existence of the resided secret monitoring program through task manager. Finally, the combination of KFsensor and honeywall (ROO) is implemented as a honeynet.

Inside the honeypot, there are deifferent types of database, some are nonsensitive such as publication, announcements, and news and some are classified such as personal information, research project and datum, and information of important partnership. As for classified database, access area are authorized according to the employee's responsibilities. For example, as for classified research projects, the accounting department only can access those authorized field such as the name or index of research projects and its corresponding budget and real expanses of a table that store in the database. The detailed technique information of the research project is encrypted and only authorized users can access it. To test the validation of the developed honeynet, some application softwares are adopted such as IBM App Scan, Brutus, Xscan, Fluxay 5.0, CurrPorts, Wireshark, and etc. The simulated network is established based on VMware techniques at the laboratory of Department of Computer Science in the Chinese Culture University.

Different type of access or intrusion, the number of packets is different. Thus the attacking packets collected for each day are also different. To collect and statistic the packet information, Wireshark software has been adopted for packet sniffing as shown in Figure 8. All of the collected packets are stored in the database of Management Server. Beside, to avoid the record cleaning or spoofing from hackers, all of the system logs and security logs are timely copied and stored with hidden mode in another database by developed hidden resided application program. Therefore, even the hackers try to clean their record of activity; we still possess the original completely record and information for comparison and analysis.

Some different attacks such as probe, DoS (Denial of Service), DDoS (Distributed DoS), R2L (remote to local), U2R (User to Root), evading IDS and etc. as tabulated in Table 4 are experimented to measure the performance of Snort IDS and the proposed IDS mechanism.

Table 4. Number of different attacks provided for test

Attacking type	Numbers	Rate
Probe	1,200	~11.5%
DoS	600	
DDoS	1,820	
R2L	1,220	
U2R	1,520	
Evading IDS	1,620	
Normal access	＞60,000	~88.5%

In which the probe and DoS attack are performed automatically. As to the other attacking approaches such as DDoS, R2L, U2R, and evading IDS, each of 20 different special attacks are designed and applied for intrusive experiment. Among those attacking, the distributed DoS and evading IDS attacks are highly emphasis for testing.

The network hacking is initiated through preparation, initiation, and cleaning steps such as ports or vulnerabilities scanning, sniffing, password cracking, privilege escalating, tunnel establishing, zombie arrangement, and track cleaning as shown in Figure 5~9.

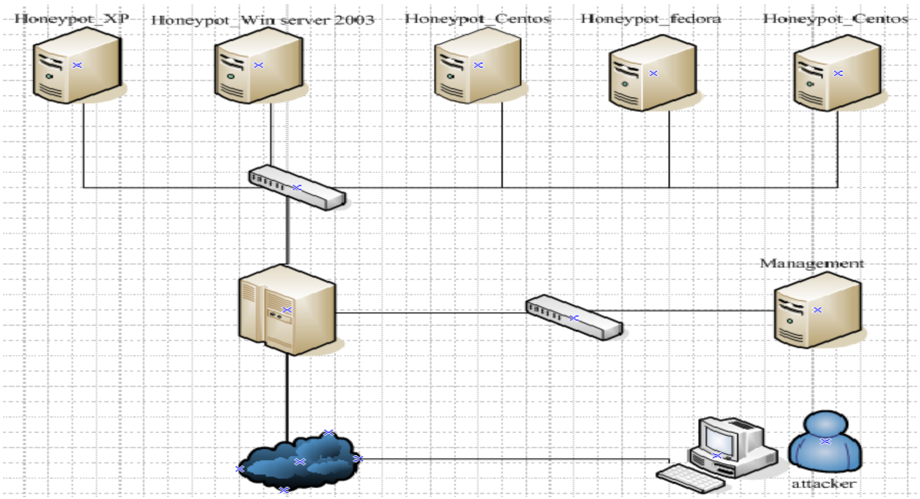
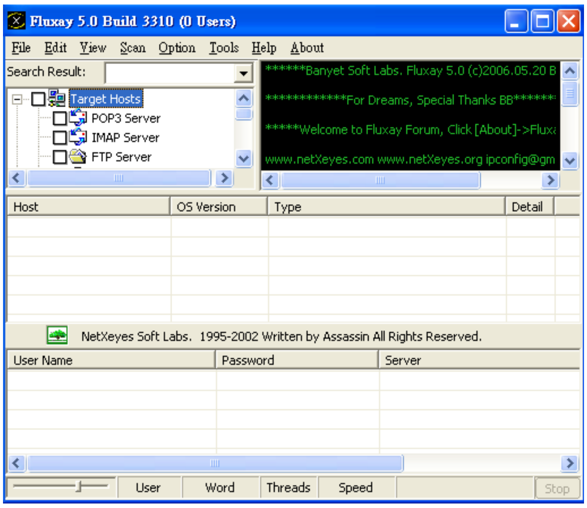


Fig. 4 The architecture of implemented honeynet.



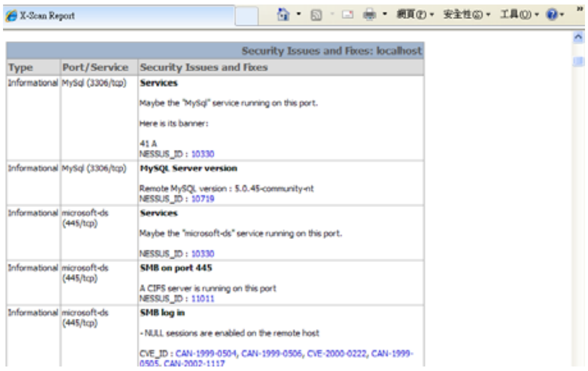
```
******Banyet Soft Labs. Fluxay 5.0 (c)2006.05.20 B
************For Dreams, Special Thanks BB******
*****Welcome to Fluxay Forum, Click [About]->Fluxa
www.netXeyes.com www.netXeyes.org ipconfig@gm
```

Host	OS Version	Type	Detail

NetXeyes Soft Labs. 1995-2002 Written by Assassin All Rights Reserved.

User Name	Password	Server

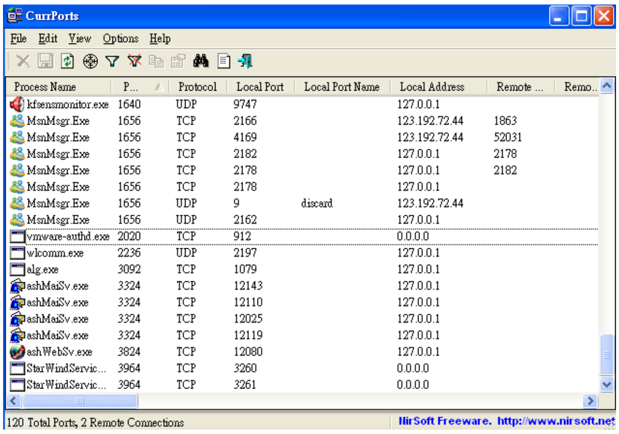
Fig. 5 Demonstration of vulnerability scanning by Fluxay 5.0.



Security Issues and Fixes: localhost

Type	Port/Service	Security Issues and Fixes
Informational	MySql (3306/tcp)	**Services** Maybe the "MySql" service running on this port. Here is its banner: 41 A NESSUS_ID : 10330
Informational	MySql (3306/tcp)	**MySQL Server version** Remote MySQL version : 5.0.45-community-nt NESSUS_ID : 10719
Informational	microsoft-ds (445/tcp)	**Services** Maybe the "microsoft-ds" service running on this port. NESSUS_ID : 10330
Informational	microsoft-ds (445/tcp)	**SMB on port 445** A CIFS server is running on this port NESSUS_ID : 11011
Informational	microsoft-ds (445/tcp)	**SMB log in** - NULL sessions are enabled on the remote host CVE_ID : CAN-1999-0504, CAN-1999-0506, CVE-2000-0222, CAN-1999-0505, CAN-2002-1117

Fig. 6 Statistical results of vulnerability scanning.



Process Name	P...	Protocol	Local Port	Local Port Name	Local Address	Remote ...	Remo..
kfsensmonitor.exe	1640	UDP	9747		127.0.0.1		
MsnMsgr.Exe	1656	TCP	2166		123.192.72.44	1863	
MsnMsgr.Exe	1656	TCP	4169		123.192.72.44	52031	
MsnMsgr.Exe	1656	TCP	2182		127.0.0.1	2178	
MsnMsgr.Exe	1656	TCP	2178		127.0.0.1	2182	
MsnMsgr.Exe	1656	TCP	2178		127.0.0.1		
MsnMsgr.Exe	1656	UDP	9	discard	123.192.72.44		
MsnMsgr.Exe	1656	UDP	2162		127.0.0.1		
vmware-authd.exe	2020	TCP	912		0.0.0.0		
wlcomm.exe	2236	UDP	2197		127.0.0.1		
alg.exe	3092	TCP	1079		127.0.0.1		
ashMaiSv.exe	3324	TCP	12143		127.0.0.1		
ashMaiSv.exe	3324	TCP	12110		127.0.0.1		
ashMaiSv.exe	3324	TCP	12025		127.0.0.1		
ashMaiSv.exe	3324	TCP	12119		127.0.0.1		
ashWebSv.exe	3824	TCP	12080		127.0.0.1		
StarWindServic...	3964	TCP	3260		0.0.0.0		
StarWindServic...	3964	TCP	3261		0.0.0.0		

120 Total Ports, 2 Remote Connections   NirSoft Freeware. http://www.nirsoft.net

Fig. 7 Monitoring the network traffic by CurrPorts.

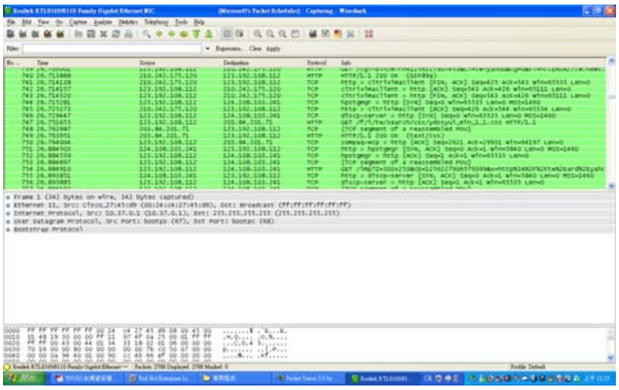
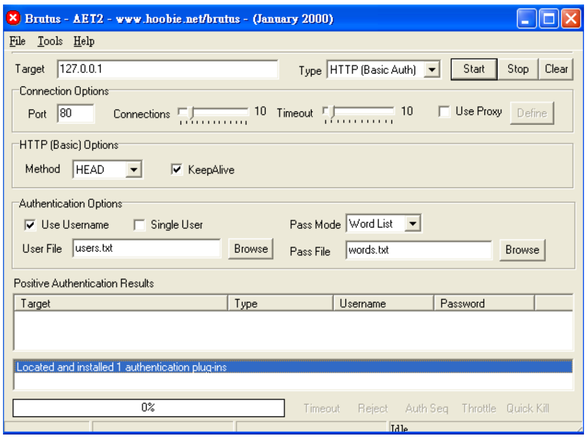


Fig. 8 Demonstration of networking traffic sniffing by Wireshark software.



Brutus - AET2 - www.hoobie.net/brutus - (January 2000)

Target 127.0.0.1　Type HTTP (Basic Auth)　Start　Stop　Clear

Connection Options
Port 80　Connections 10　Timeout 10　□ Use Proxy　Define

HTTP (Basic) Options
Method HEAD　☑ KeepAlive

Authentication Options
☑ Use Username　□ Single User　Pass Mode Word List
User File users.txt　Browse　Pass File words.txt　Browse

Positive Authentication Results

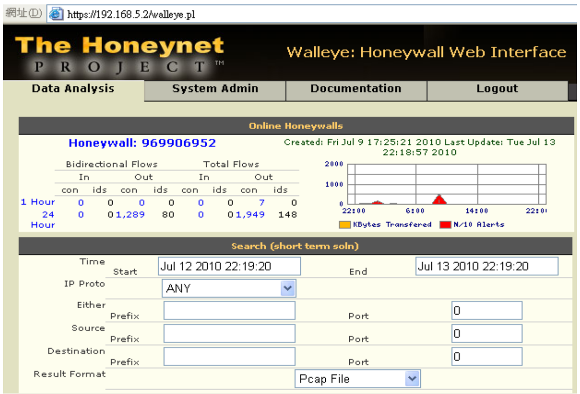
Target	Type	Username	Password

Located and installed 1 authentication plug-ins

0%　Timeout　Reject　Auth Seq　Throttle　Quick Kill

Idle

Fig. 9 Demonstration of password cracking.



網址(D) https://192.168.5.2/walleye.pl

Online Honeywalls

Honeywall: 969906952　Created: Fri Jul 9 17:25:21 2010 Last Update: Tue Jul 13 22:18:57 2010

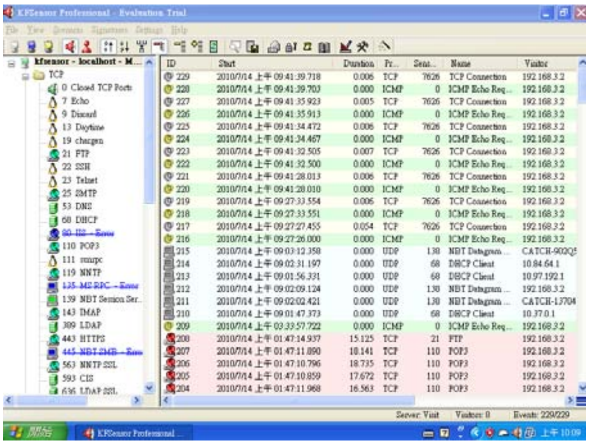
	Bidirectional Flows				Total Flows			
	In		Out		In		Out	
	con	ids	con	ids	con	ids	con	ids
1 Hour	0	0	0	0	0	0	7	0
24 Hour	0	0	1,289	80	0	0	1,949	148

Search (short term soln)

Time Start Jul 12 2010 22:19:20　End Jul 13 2010 22:19:20
IP Proto ANY
Either Prefix　Port 0
Source Prefix　Port 0
Destination Prefix　Port 0
Result Format Pcap File

Fig. 10 Demonstration of attacking detection by honeynet.

Shown in Figure 10 is the detection of X-scan attacking by the deployed honeynet. Shown in Figure 11, that is abnormal behavior initiated from the attacking host which was detected by KFsensor.



KFSensor Professional - Evaluation Trial

kfsensor - localhost - M...
TCP
0 Closed TCP Ports
7 Echo
9 Discard
13 Daytime
19 chargen
21 FTP
22 SSH
23 Telnet
25 SMTP
53 DNS
68 DHCP
~~80 IIS - Error~~
110 POP3
111 rmnrpc
119 NNTP
~~135 MS RPC - Error~~
139 NBT Session Ser...
143 IMAP
389 LDAP
443 HTTPS
~~445 NBT SMB - Erro~~
563 NNTP SSL
593 CIS
636 LDAP SSL

ID	Start	Duration	Pr...	Sent...	Name	Visitor
229	2010/7/14 上午 09:41:39.718	0.006	TCP	7626	TCP Connection	192.168.3.2
228	2010/7/14 上午 09:41:39.703	0.000	ICMP	0	ICMP Echo Req...	192.168.3.2
227	2010/7/14 上午 09:41:35.923	0.005	TCP	7626	TCP Connection	192.168.3.2
226	2010/7/14 上午 09:41:35.913	0.000	ICMP	0	ICMP Echo Req...	192.168.3.2
225	2010/7/14 上午 09:41:34.472	0.006	TCP	7626	TCP Connection	192.168.3.2
224	2010/7/14 上午 09:41:34.467	0.000	ICMP	0	ICMP Echo Req...	192.168.3.2
223	2010/7/14 上午 09:41:32.505	0.007	TCP	7626	TCP Connection	192.168.3.2
222	2010/7/14 上午 09:41:32.500	0.000	ICMP	0	ICMP Echo Req...	192.168.3.2
221	2010/7/14 上午 09:41:28.013	0.006	TCP	7626	TCP Connection	192.168.3.2
220	2010/7/14 上午 09:41:28.010	0.000	ICMP	0	ICMP Echo Req...	192.168.3.2
219	2010/7/14 上午 09:27:33.554	0.006	TCP	7626	TCP Connection	192.168.3.2
218	2010/7/14 上午 09:27:33.551	0.000	ICMP	0	ICMP Echo Req...	192.168.3.2
217	2010/7/14 上午 09:27:27.455	0.054	TCP	7626	TCP Connection	192.168.3.2
216	2010/7/14 上午 09:27:26.000	0.000	ICMP	0	ICMP Echo Req...	192.168.3.2
215	2010/7/14 上午 09:03:12.358	0.000	UDP	138	NBT Datagram ...	CATCH-902Q5
214	2010/7/14 上午 09:02:31.197	0.000	UDP	68	DHCP Client	10.84.64.1
213	2010/7/14 上午 09:01:56.331	0.000	UDP	68	DHCP Client	10.97.192.1
212	2010/7/14 上午 09:02:09.124	0.000	UDP	138	NBT Datagram ...	192.168.3.2
211	2010/7/14 上午 09:02:02.421	0.000	UDP	138	NBT Datagram ...	CATCH-13704
210	2010/7/14 上午 09:01:47.373	0.000	UDP	68	DHCP Client	10.37.0.1
209	2010/7/14 上午 03:33:57.722	0.000	ICMP	0	ICMP Echo Req...	192.168.3.2
208	2010/7/14 上午 01:47:14.937	15.125	TCP	21	FTP	192.168.3.2
207	2010/7/14 上午 01:47:11.090	10.141	TCP	110	POP3	192.168.3.2
206	2010/7/14 上午 01:47:10.796	18.735	TCP	110	POP3	192.168.3.2
205	2010/7/14 上午 01:47:10.859	17.672	TCP	110	POP3	192.168.3.2
204	2010/7/14 上午 01:47:11.968	16.563	TCP	110	POP3	192.168.3.2

Server: Visit　Visitors: 0　Events: 229/229

Fig. 11 Demonstration of the detection of abnormal behavior by KFsensor.

There are some statistical results of different types of information security event that has been collected within a whole month as tabulated in Table 5. From the table, one can easily find out that SQL attacking does possess significant ratio and reach more than 40%. The result is consistent to the announcement of 15 important information security attacking strategies that pointed out by Verizon Business [16]. In their report, keylogger and spyware, trapdoor and command/control program, and SQL Injection are the top 3 on the list that induce significant problem of information leakage.

Table 5. Statistical results for some information security event within one month

Item	Type of Information Security Event		No. of Attempts	Rate%
1	Login and Password	FTP(login_failed)	768	16.27
		MS_SQL_Srver(null)	1919	40.67
		MS_SQL(empty)	1921	40.71
2	Authentication	VNC	27	0.57
3		Realvnc_Server(bypass)	21	0.45
4		MS_SQL(overflow)	11	0.23
5	SQL_Injection	Var_Char	2	0.04
6		Select	2	0.04
7		Declare	1	0.02
8		Create_table	1	0.02
9	Dos	Working_resources (http)	4	0.08
10		IPswitch_wsftp_rest (large argument)	1	0.02
11		Apache_modssl_custom (error document)	1	0.02
12	MS_SQL	Server_XP (command execution)	1	0.02
13		Packet type(unknown)	11	0.23
14	Others	HTTP Proxy	10	0.21
15		Echo Message(Superscan)	7	0.15

Table 6. The detection comparison of general Intrusion between proposed and original IDS

Kind	Snort IDS			Proposed IDS		
Condition	*Normal*	*Misuse*	*Anomaly*	*Normal*	*Misuse*	*Anomaly*
Detection rate	96%	82%	54%	98%	92%	86%

The detection performance between Snort IDS and the proposed IDS based on without utilizing distributed strategy for attacking is compared and shown as tabulated in Table 6. In which the performance of the proposed IDS mechanism is a little better than Snort IDS.

From Table 6, we can find out that the successful detection rate of Snort IDS for abnormal intrusive is about 54% and the misuse detection rate is 82%. As to the proposed IDS mechanism, the successful detection rate for abnormal intrusive detection is higher and reaches 86% and the misuse detection rate is about 92%. Both of the Snort IDS and the proposed IDS mechanism possess low false positive and false negative rate. That means the IDSs can easily and successfully identify those false signal.

Since the distributed and organized attacking are highly increased, some distributed attacks are emphasized and provided for measuring the performance of Snort IDS and the proposed IDS in order to provide a well handle process experience. The experimental results are shown in Table 7.

Table 7. Statistical result of intrusion detection from distributed and organized attacking by proposed and Snort IDS

Kind	Snort IDS	Proposed IDS
Correct Detection	36.25%	82.92%

From Table 7, we can easily find out that the performance is differed explicitly between Snort IDS and the proposed IDS mechanism.

From the tabulated result, the successful detection rate for Snort IDS to detect the distributed and organized attack is about 36.25%.

As for the proposed IDS mechanism, the successful detection rate to detect distributed and organized attack has significant improved to reach about 82.92%. The above demonstration has revealed the advantage of adopting the multi-dimensional hidden Markov model for intrusion analysis. It is explicit that the proposed IDS mechanism does perform better than Snort IDS. That is because the dynamical rationale such as information updating in realtime/timely has been adopted in the proposed mechanism. Moreover, adopting the rationale of HMM and ant colony algorithm for intrusive analysis does provide more accurate analyzed results.

Ⅳ. CONCLUSION

In this paper, a novel IDS mechanism is proposed. In the proposed mechanism, there are three modules including the monitor module, track module, and analysis module developed. All of the intrusive information and audit data are collected and integrated for analysis based on multidimensional hidden Markov model combined with autocorrelation and cross-correlation measurement to provide higher accuracy and efficient intrusion detection performance. After then, the fuzzy inferring rule is applied for intrusion recognition and identification. The console module is assigned to manage the performance of the system, control all of the sensors for monitoring security events and generate alerts and offer periodically reports and present proposals for taking suitable response and making optimal decision. In addition, the proposed IDS mechanism does possess dynamic characteristic. In which if there is any new virus or attacking strategy occurred, the new signature will immediately be added to the database belong to the Expert module of the IDS for record updating even though the system is under operation. Thus, to achieve the goal of keeping the proposed IDS always performed under optimal condition.

Experimental results demonstrate that the proposed IDS mechanism does achieve the goal of exactly, correctly, effectively, and completely collected and analyzed all of the normal and abnormal intrusive information. The proposed IDS mechanism could exactly indicate the kernel cause of information security events and provided the information for remedy those possible leakages in timely. Especially, the adoption of dynamically and timely information updating and the application of HMM combined with ant colony algorithm for intrusion analysis does gain more advantages than Snort IDS. The proposed mechanism can be applied for perspective application on information security.

On October 2009, the Gartner Inc. released the IT evolution of 2010, in which cloud computing has become the most significant issue of the 10 top IT issues [16]. In Taiwan, the government also has revealed 8 important items regarding the infrastructure construction for the application of cloud computing. As regarding the information security of cloud computing, those issues related with the application and service of cloud computing must be deeply concerned such as the equipment security of the client side, the threats of web site and webpage, the detection and diagnosis and surveillance of intrusion, the access and security of database on the cloud side, the detection of system leakage and the monitor of real-time repairing process, the management of server system, the management of mobile e-commerce processing, and the integrated analysis of associated security information and issues. In addition, different types of network hacking will generate different information security threats and might provide different rationale of protection strategies. Nevertheless, currently the most obsessive network security problem is insider threats. Therefore, effective responses to mitigate the harm from Internet hacking is a must. In the future, some advanced research such as the information security issues of cloud computing and the designation of most appropriate strategies to mitigate from intruder and insider threat will be implemented.

ACKNOWLEDGEMENTS

The research of this paper is sponsored by Chinese Culture University of Taiwan, R.O.C.

REFERENCES

[1] Gupta, K. K, Nath, B., and Kotagiri, R.,“Layered Approach Using Conditional Random Fields for Intrusion Detection,” IEEE Transactions on Dependable and Secure Computing, Vol. 7, No. 1, pp. 35-49, Jan 2010.

[2] Mohammad, G. G., Reza, M. and Hadi, S. Y., “Intrusion Detection by New Data Description Method,” 2010 International Conference on Intelligent Systems, Modeling and Simulation, pp. 1-5, 2010.

[3] Rabiner, L. R., Juang, B. H., Levinson, S. E., and Sondhi, M. M., “Some properties of continuous hidden Markov model representations,” AT&T Tech. Journal, Vol. 64, No. 6, pp. 1251-1270, 1985.

[4] Depren, O., Topallar, M., Anarim, E., and Ciliz, M. K., “An intelligent intrusion detection system (IDS) for anomaly and misuse detection in computer networks,” Expert Systems with Application, Vol. 29, pp. 713-722, 2005.

[5] Tsai, C. F., Hsu, Y. F., Lin, C. Y., and Lin, W. Y., “Intrusion detection by machine learning: A riview,” Expert Systems with Application, Vol. 36, pp. 11994-12000, 2009.

[6] Ajith, A., Ravi, A., J. and Johnson T., “D-SCIDS: Distributed soft computing intrusion detection system,” Journal of Network and Computer Applications, Vol. 30, pp. 81–98, 2007.

[7] Gabriel, R., Hoppe, T., Pastwa, A. and Sowa, S., “Analyzing Malware Log Data to Support Information and Event Management: Some Research Results,” First International Conference on Advances in Databases, Knowledge, and Data Applications, pp. 108-113, 2009.

[8] Pfleeger, S. L., Predd, J. B., Hunker, J. and Bulford, C., “Insiders Behaving Badly: Addressing Bad Actors and Their Actions,” IEEE Transactions on Information Forensics and Security, Vol. 5, No. 1, pp. 169-179, March 2010.

[9] Joshua O. N., “Log Analyzer for Network Forensics and Incident Reporting,” 2010 International Conference on Intelligent Systems, Modeling and Simulation, pp.356-361, 2010.

[10] Renuka, P. B. and Abraham, A., “Hybrid Framework for Behavioral Prediction of Network Attacking Using Honeypot and Dynamic Rule Creation with Different Context for Dynamic Blacklisting,” Proceedings of 2010 Second International Conference on Communication Software and Networks, pp. 471-476, Singapore , Feb 26-28, 2010.

[11] Lee, D. H., Kim, D. Y., and Jung, J. I., “Multi-Stage Intrusion Detection System Using Hidden Markov Model Algorithm,” Proceedings of IEEE International Conference on Information Science and Security, pp. 72-77, 2008.

[12] Praveen, K., Rao, K., and Kamakshi, P.,

"Ant-Colony Optimization Algorithm for Computer Intrusion Detection," International Advanced Database Conference 2006, San Diego, U.S.A., June 27~29, 2006.

[13] Accessed on June 20, 2010, http://en.wikipedia.org/wiki/Ant_colony_algorithm

[14] Ahmad, I., Abdullah, A. B., and Alghamdi, A. S., "Comparative Analysis of Intrusion Detection Approaches," 12th International Conference on Computer Modelling and Simulation, pp. 586-591, Cambridge, United Kingdom , March 24~26, 2010.

[15] Li, Y. Z., Wang, R. S., Xu, J., Yang, G. and Zhao, B., "Intrusion Detection Method Based on Fuzzy Hidden Markov Model, " 6th International Conference on Fuzzy Systems and Knowledge Discovery, pp.470-474, China, August 14~16, 2009.

[16] Verizon Business, Accessed on Sep 2010, http://www.verizonbusiness.com/

[17] Accessed on Sep 2009 and Apr 2010, Gartner Incorporation, http://www.gartner.com/