A Model for Intrusion Detection Based on Undefined Distance

Ram Kumar, Sarvesh Kumar, Kolte V.S

Abstract-In this paper, we introduced the intrusion detection system and the uncertain theory, and point out two important prerequisite that the IDS work normally must depend on, and in view of the prerequisite, the paper proposed a solution which is based on uncertain distance and the active defense technology anti-host intrusion. The solution can distinguish normal event from the unknown event efficiently, and can detect unknown event. This paper proposed the active defense technology anti-host intrusion based on uncertain distance. The system can not only judge normal event, but also can detect unknown event. The system can judge whether an event is harmful, and can store the eigenvector of suspicious event to "normal event set" or "intrusion event set" automatically.

Keywords-Intrusion event; Active Defense; Uncertain theory; Uncertain distance; Intrusion Detection; Intrusion Event; Anti-Host; Intrusion Event Set;

I. INTRODUCTION

In computer there are some security vulnerabilities through which computer virus and attackers intrude a computer system (operating system, application software and hardware equipments) without owner's knowledge.

As per concern of network security technology developed to resistance to network attack's and experienced the development process in two ways, i.e. static and dynamic, from passive defense to active defense, from the local defense to overall defense [1]

Intrusion prevention measures, such as encryption and authentication, can be used in ad-hoc networks to reduce intrusions, but cannot eliminate them. For example, encryption and authentication cannot defend against compromised mobile nodes, which often carry the private keys. Integrity validation using redundant information (from different nodes), such as those being used in secure routing, also relies on the trustworthiness of other nodes, which could likewise be a weak link for sophisticated attacks. To secure mobile computing applications, we need to deploy intrusion detection and response techniques, and further research is necessary to adapt these techniques to the new environment, from their original applications in fixed wired network.

The intrusion behavior and the legal act are may differentiate, that is to say, may judge the nature of this behavior through the extraction behavior pattern characteristic. An intrusion detection system needs to solve two problems [2]

a. How to extract the data of behavioral characteristics entirely and reliably;
b. How to determine the nature of the act by the characteristic data efficiently and accurately.

The purpose of intrusion detection is to identify internal and external conduct which is exceed its authority, misuse and abuse, and at the same time to protect the legitimate users to make use of the resources of system efficiently.

Host-based intrusion detection system is installed in the host need to focus on detection, monitors and analyzes the host audit records. If it is found that the activities of objects are very suspicious, intrusion detection system will take corresponding measures. Host Intrusion Detection System commonly used in the analysis of "possible attacks", can provide more detailed information on the evidence that the intruders tried to implement a "dangerous order", to distinguish the specific acts of the intruders.

In this paper, we use the uncertain theory to distinguish between normal events and intrusion events. In this model we combine the two deferent methods and its positive features of active defence technology and the uncertain theory.

II.PRELIMINARIES

A. Uncertainty Space

Let Γ be a nonempty set, and let τ be a σ –algebra over Γ Each element $\Lambda 0 \tau$ is called an event. In order to present an axiomatic definition of uncertain measure, it is necessary to assign to each event Λ a number $M\ \{\Lambda\}$ which indicates the level that Λ will occur. In order to ensure that the number $M\ \{\Lambda\}$ has certain mathematical properties, Liu proposed the following four axioms:

Axiom 1. (Normality) $M\ \{\ \Gamma\ \} = 1$.

Axiom 2. (Monotonicity) $M\ \{\ \Lambda_1\} \leq M\ \{\ \Lambda_2\ \}$ whenever $\Lambda\ 1\ \Lambda 2$

Axiom 3. (Self-Duality) $M\{\ \Lambda\ \} + M\{\ \Lambda^c\ \} = 1$ for any event Λ

Axiom 4. (Countable Subadditivity) For every countable sequence of events $\{\Lambda_i\}$, we have

Manuscript received July 09, 2011.

Ram Kumar is with Computer Science & Engineering, MSS"S College of Engineering & Technology, Jalna, India, Mobile No.+918698582404, (e-mail: hr.coet@gamil.com).

Sarvesh Kumar is with Computer Science & Engineering , Rama Institute of Engineering & Technology, Kanpur (U.P), India , Phone/ Mobile No.+917398348072, (e-mail: er.sarvesh05@gmail.com).

Kolte V.S. is with Computer Science & Engineering, MSS"S College of Engineering & Technology, Jalna, India, Mobile No.+918275321854, (e-mail: koltevs@gamil.com).

$$\text{M}\ \{\cup_{i=1}^{\infty}\Lambda_i\} \leq \sum_{i=1}^{\infty} M\ \{\Lambda_i\} \qquad (1)$$

B. Uncertain Variables

Definition: An uncertain variable is a measurable function ξ from an uncertainty space (Γ, τ, M) to the set of real numbers, i.e., for any Boral set B of real numbers, the set

$$\{\xi\epsilon\text{B}\} = \{\gamma\epsilon\Gamma|\xi(\gamma)\epsilon\text{B}\} \qquad (2)$$

is an event.

Definition: An n-dimensional uncertain vector is a measurable function from an uncertainty space (Γ, τ, M) to the set of n-dimensional real vector, i.e., for any Boral set B of $\mathfrak{R}^n$, the set

$$\{\xi\epsilon\text{B}\}=\{\gamma\epsilon\Gamma|\xi(\gamma)\epsilon\text{B}\} \qquad (3)$$

is an event.

C. Identification Function

Definition: an uncertain variable ξ is said to be have a first identification function λ if $\lambda\ (x)$ is a nonnegative function on $\mathfrak{R}^n$ such that

$$\sup(\lambda(x) + \lambda(y)) = 1; \qquad (4)$$

For any set B for real number, we have

$$\text{M}\{\xi\epsilon\text{B}\} = \begin{cases} \sup \lambda_{x\epsilon\text{B}}(x), & \text{If } sup\ \lambda_{x\epsilon\text{B}} < 0.5 \\ 1\text{-}\sup \lambda_{x\epsilon\text{B}^c}(x), & \text{If } sup\ \lambda_{x\epsilon\text{B}} < 0.5 \end{cases} \qquad (5)$$

Example: By a triangular uncertain variable we mean the uncertain variable fully determined by the triplet (a,b,c) of crisp numbers with a<b<c, whose first identification function

$$\lambda(x) = \begin{cases} \frac{x-a}{2(b-a)}, & \text{if } a \leq x \leq b \\ \frac{x-c}{2(b-c)}, & \text{if } b \leq x \leq c \end{cases} \qquad (6)$$

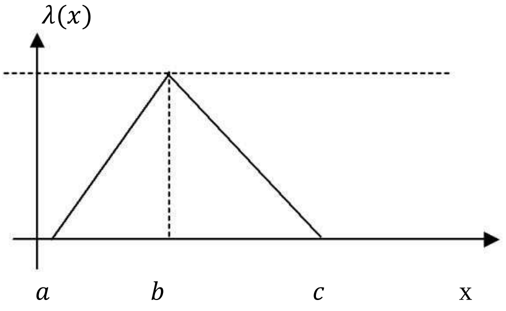


Figure1. First Identification function λ

D. Expected Value

Definition: let ξ be an uncertain variable. Then the expected value of ξ is defined by

$$\text{E}[\xi] = \int_0^{\infty} \text{M}\{\xi \geq r\}dr - \int_0^{\infty} \text{M}\{\xi \leq r\}dr \qquad (7)$$

provide that at least one of the two integrals is finite

E. Distance

Definition: the distance between two uncertain variables ξ and η is defines as

$$d(\xi,\eta) = \text{E}[|\xi - \eta|] \qquad (8)$$

Example: let $\xi = (a_1, b_1, c_1)$ and $\eta = (a_2, b_2, c_2)$ be triangular uncertain variable such that

$(a_1, c_1)\cap(a_2, c_2)=\emptyset$. Than

$$d(\xi,\eta) = \frac{1}{4}(|a_1\text{-}c_2|+2|b_1\text{-}b_2|+|c_1\text{-}c_2|) \qquad (9)$$

III. OVERALL SYSTEM STRUCTURE

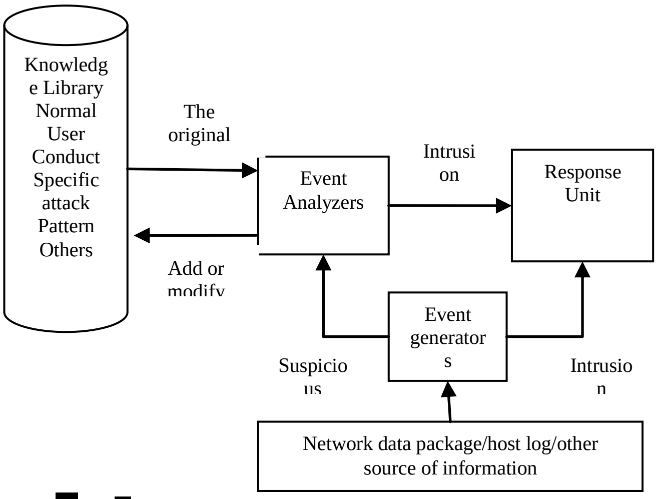
A proposed common Intrusion Detection Framework, which is showed in Figure 2[1] [4]

In order to explain the design and the framework of active defence system against host intrusion illustrate using Figure 2. In this proposed system that consists three sets:

a. “Normal event set” N, ”
b. “Intrusion event set” I and
c. “Suspicious event set” S.

The normal event signatures are stored in N, the known intrusion event signatures are stored in I, and the suspicious event signatures are stored in S, Let U is a set that represents full set, and the relationship between the various sets is [4]

U= N∪ S ∪ I
N∩ I= ∅
N∩S= ∅


Knowledge Library Normal User Conduct Specific attack Pattern Others
The original
Add or modify
Event Analyzers
Intrusi on
Response Unit
Event generator s
Suspicio us
Intrusio n
Network data package/host log/other source of information

S∩I= ∅

Figure 2 the common intrusion detection framework

The information of the three sets can be obtained through the following method. The flow chart of active defence anti-host intrusion system is showed in Figure 3[4].

A. *The establishment of “normal event set”*

For the establishment of “normal event set” we have to setting up a local area network environment that is free from the virus and malicious procedure , now we get the information which generated when the network exchange information and then encrypt it though MD5 then save it to "normal event set"

B. The establishment o*f "intrusion event set"*

For the setting up the "intrusion event set" we pick-up the signature of the known virus and malicious procedure using signature extraction software, and encrypt through MD5 then save it to "intrusion event set"

C. The establishment *of "suspicious event set"*

The establishment of suspicious event set, we monitor the process intercepts and capture all events and suspect it for a while. Then pick-up the signature of the event using signature extraction software and encrypt through MD5, then compare it with the "normal event set", if the ongoing event is legal event, then let it continue, if not then compare its signature in "intrusion event set", if it seems to harmful then kill it, else suspend the process and save it to "suspicious even set".

D. How to deal with the suspicious process

Whether to allow the event whose signature was saved to "suspicious event set" to run depends on the detection by the uncertainty distance.

a) The suspicious operation sequence

In order to compute the uncertainty distance, we must give the uncertainty of each characteristic of the process.

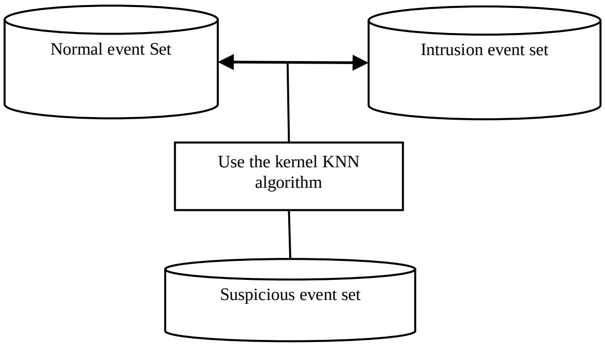


Figure 3 the framework of active defiance system

After the analysis of the intrusion event, we can sum up several common characteristics and uncertainty as follows [4].

a. Abnormal file access. When the harmful process infects other executable files, it is generally to traverse all the executable files, which is not the characteristic of common files, and if finding a process with the kind of characteristic, we can recognize it is a suspicious process. Its uncertainty is in the range of 0.2~0.4, based on triangular uncertain variable, we can set a=0.2, c=0.4.

b. Abnormal memory operation. When the Trojan horse infects the common files and runs, it will erase or move or replace the memory, and the operation is the unique to the virus generally. Its uncertainty is in the range of 0.15~0.3, based on triangular uncertain variable, we can set a=0.15, c=0.3.

c. The leaking secrets operation. In order to steal information from users, Trojan horse will collect users' information automatically. This is an important feature of Trojan horse. Its uncertainty is in the range of 0.1~0.3, based on triangular uncertain variable, we can set a=0.1, c=0.3.

d. Self-replication operation. If a process has the self-replication code sequence, we should suspect it is Trojan horse. Its uncertainty is in the range of 0.05~0.25, based on triangular uncertain variable, we can set a=0.05, c=0.25.

e. The call of private API function. In order to reduce its volume, Trojan horse often calls a large number of API functions, including private API function. The process with this feature can be classified as suspicious ones. Its uncertainty is in the range of 0.25~0.45, based on triangular uncertain variable, we can set a=0.25, c=0.45.

f. Presence in memory. The Trojan horse must be presented in memory for a long-term in order to achieve the purpose of infection and spreading. So a process which automatically resides in memory for a long-term can be considered as a harmful process. Its uncertainty is in the range of 0.3~0.48, based on triangular uncertain variable, we can set a=0.3, c=0.48.

3.4.2 The establishment of model

Based on the analysis above, now we have 6 evaluation factors to consist of an assessment program

$$x = \left(t_1(x), t_2(x), \dots\dots\dots\dots t_6(x)\right), 0 \leq t_i(x) \leq 0.5, 1 \leq i \leq 6$$

Through questionnaire by experts, we can obtain the weight of each factor above as the follows:

$$\breve{w}_j (j = 1,2, \dots\dots\dots,6), and\ \sum_{j=1}^{6} \breve{w}_j=1, (\breve{w}_j \geq 0)$$

Based on the above analysis and the First Identification Function, in Windows2000 Server environment, we selected 40 PE files of system as a security database, taken another 20 PE file of system as the samples of classification. According to the virus signature provided by Kingsoft, we obtained 30 Trojan horse files through various channels, of which 10 were randomly selected from the virus database, and 20 samples as the classification samples. Based on the above 6 characteristic, we tested the 90 files, according to the triangular uncertain variable; we can take their expected value as the $b_i (1 \leq i \leq 6)$ value. Then the triangular uncertain variable $\xi = (a + b + c)$ has an expected value

$$\mathrm{E}[\xi] = \left(\sum_{i=1}^{6} \breve{w}_i \,(a_i + 2b_i + c_i) \right) /4$$

According to the above 40 PE files, we can obtain the security set(S)'s expected value as the follows:

$$\mathrm{E}[S] = \left\{ \left(\sum_{i=1}^{40} \breve{w}_i \,(a_i + 2b_i + c_i) \right) /4 \right\} /40$$

According to the above 30 virus files, we can obtain the virus set (V)'s expected value as the follows:

$$\mathrm{E}[V] = \left\{ \left(\sum_{i=1}^{20} \breve{w}_i \,(a_i + 2b_i + c_i) \right) /4 \right\} /20$$

We can get each normal event's distance to normal set (N)'s expected value, and use the longest Nl as the security threshold value.

We can get each intrusion event's distance to intrusion set(I)'s expected value, and use the longest Il as the virus threshold value.

Now, if one process is starting, we can intercept its signature $y = (t_1(y), t_2(y), \dots\dots\dots\dots t_6(y), 0 \leq t_i(y) \leq 0.5, 1 \leq i \leq 6$, and set $a_y = \sum_{i=1}^{6} \breve{w}_i\ \ a_i$, $b_y = \sum_{i=1}^{6} \breve{w}_i\ \ b_i$, $c_y =$

$\sum_{i=1}^{6} \widetilde{w_i}\, c_i$, then y can be signed (a_y, b_y, c_y), as the same , we can sign the normal set (N)'s center as (a_n, b_n, c_n) , so we can obtain the distance of y and ((V) :

$$d(y,n) = \frac{1}{4}\left(\left|a_y - c_n\right| + 2\left|b_y - b_n\right| + \left|c_y - a_n\right|\right),$$

If $d(y,n) \leq NI$,

we have considered y is an intrusion event, else we consider y is safety.

IV.CONCLUSION

In this paper we proposed the active defence technology anti-host intrusion based on undefined distance. The system can able to detect both normal and unknown event. The system can also detect whether an event is harmful, and can store the eigenvector of suspicious event to “normal event set” or “intrusion event set” automatically.

REFRENCES

[1] Yu, XD, Research on active defence technology with virus based on improved K-Nearest Neighbor Algorithm. 2008 PROCEEDINGS OF INFORMATION TECHNOLOGY AND ENVIRONMENTAL SYSTEM SCIENCES: ITESS 2008, VOL 4,2008,: 724-726.

[2] Teng Shaohua, a study on object-monitoring-based distributed and collaborative intrusion detection[D]. The degree of Doctor of Philosophy, Faculty of Electromechanical Engineering Guangdong University of Technology. (in Chinese)

[3] B. Liu, Theory and Practice of Uncertain Programming, 2nd ed., Springer-Verlag,Berlin, 2009

[4] Yu, XD, Research on Active Defence Technology with Host Intrusion Based on K-Nearest Neighbor Algorithm of Kernel . FIFTH INTERNATIONAL CONFERENCE ON INFORMATION ASSURANCE AND SECURITY, VOL 1, PROCEEDINGS,2009,: 411-414

[5] DENG Lujuan, LIU Tao, GAN Yong, XIONG Kun. Active Defence Technology with Virus Based on Differentiation and Hiding Process Computer Engineering, March 2007. Vol.33 No.5.117-119. (in Chinese)