Triples Anomaly Detection Security Model Based on Decision Tree

Liangcheng Lin[1, a], Song Qing[2, b], Ting Jiang[3, c] and Leiyue Zhou[3, d]

[1]China Power Information Technology of Beijing, Beijing 100085, China;

[2] Xinjiang Electric Power Company, Xinjiang 830018, China.

[3] School of of Control and Computer Science Engineering, North China Electric Power University , Beijing 102206, China.

[a]xuyonggang@sgitg.sgcc.com.cn, [b]qingsong@xj.sgcc.com.cn, [c]18046504036@163.com, [d]zlyddygyx1102@163.com

Keywords: Data-mining, Network security, triples model.

Abstract. A triples anomaly detection security model based on decision tree algorithm is designed for solving the issues--exist in the current network security detection models,such as lacking of filtering the events, and classifying the events in the network environment faintly.This model which based on the decision tree positions and distinguishes various network threats at the first step.Afterwards, dividing the degree of anomalies through analyzing the source IP address, the destination IP address and the event types of network anomalies.Proved by examples,the triples anomaly detection security model produces remarkable effects on defining the type of security incidents and determining the degree of abnormal threats.

1. Introduction

With the rapid development of Internet technology and applications, the issues of network security become more and more serious,and the network security threats are also various day by day.Emerging network security monitoring technologies, tools, products and equipments access and generate a great deal of network security incident and traffic monitoring data.Traditional manual methods and simple statistical methods have been unable to meet the need of processing the data.In order to obtain valuable information from the data, it is urgent to study a more efficient analysis and processing method[1].

At present, Internet network security system of our country is still in the embryonic stage, has not yet formed a complete industrial development chain without the core technologies to support[2]. Although the network security market showing an overall trend of rapid development,but it really lacks a unified network security system standards,and can not form a comprehensive coverage, dynamic tracking of network security pattern.The hot event model is established by literature[1],although the most frequent events can be judged by the number of abnormal events, there is no study about the source IP address and the destination IP address. The query model of intrusion system is established by literature[2], which can be issued in time, but it has not been judged after the event.Literature[11] proposed the Biba access control model, which provides a sub-level of data integrity assurance, but to a certain extent, they ignore the confidentiality.Literature[12] established the P2P network security model, which can sharply analyze the network security threat, but it is limited within the scope of P2P network, that can not explain the range of threat across networks.

Analysis of security event model established in the literature for various security incidents have both advantages and disadvantages.Weigh down, this paper established the riples anomaly detection security model fully analyzes the source, destination IP address and the event type of the unabnormal events' parameters,which can not only judge the trend of events, the trend of controlled hosts, but also the degree of abnormal.The riples anomaly detection security model filtering and sorting of data by decision tree algorithm firstly, and next puts the pick-off related parameters of the cyber threats events on the entrance detection model to determine whether the threat abnormal and what the abnormal degree.

2. Security Threats Analysis

Types of security threats become increasingly complex and difficult to prevent like Intrusion infiltration, DDoS attacks, etc. from the early simple system vulnerabilities, phishing. There is a wide range of security threats, and a variety of classifications. According to the Ministry of Industry and Information Technology issued the "Internet communications network security implementation approach", the Internet network security threats are divided into thirteen categories: Computer viruses, worms, trojans, botnets, domain name hijacking, phishing, web tampering, pages linked to horse, denial of service attacks, backdoor vulnerability, unauthorized access, spam and others[3].

As shown in Figure 1, the network system may face different sources of threats according to its own characteristics and location application. Threat sources can generally be divided into environmental information, human information and status information of the three according to their property. Environmental information refers to the physical environment of the system having been changed due to the external attack, including the utilization of memory, the utilization of CPU, the utilization of bandwidth, the utilization of hard disk and so on.Human information refers to the threats which have been caused by people's a variety of purposes and motives, including breaking into the system occasionally, malicious attacks,manufacturing viruses and malware, accessing to confidential information unauthorizedly and so on.The threats refer to those threats caused by the operating environment failure of network systems, including the importance of host, the availabilities of host, the recoverabilities of data, etc.Status information refers to the attacking events' property recording by the Snort, including the severity of the attack, the frequency of attack, the number of attacks, the types of attack and the time of attack[4].

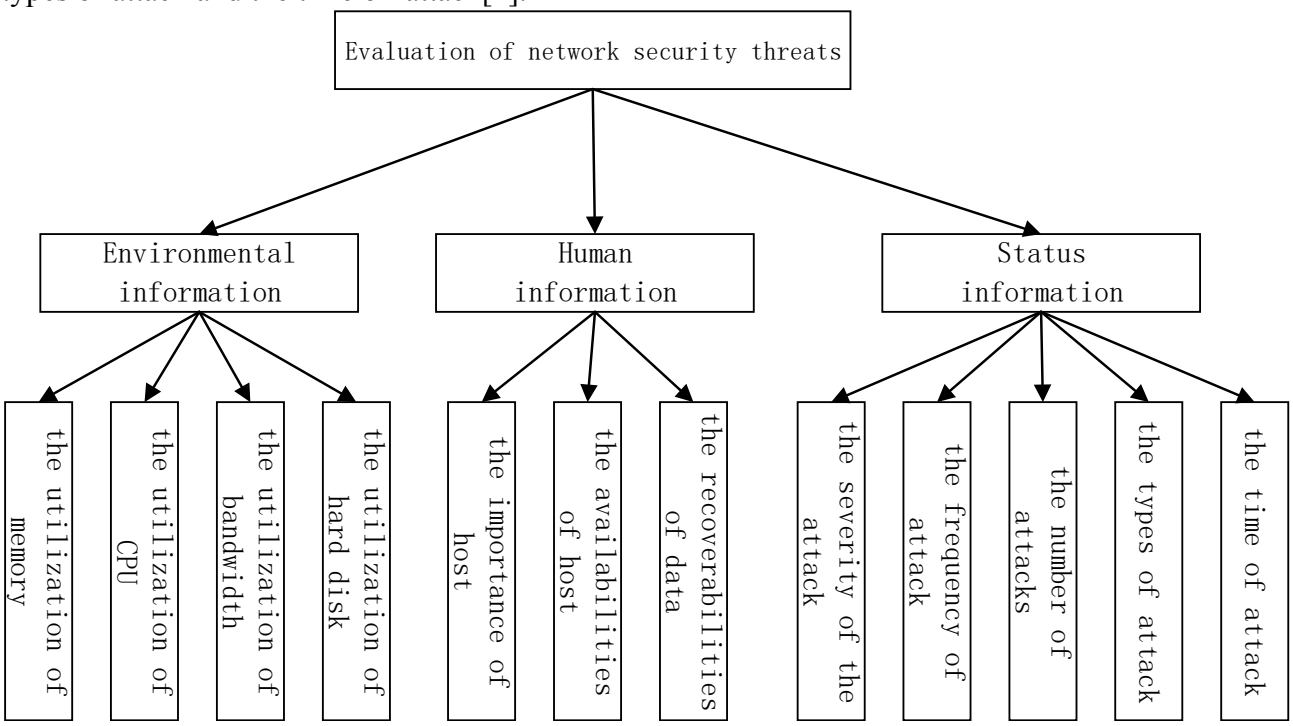

Fig. 1 Evaluation system of network security threats

When the specific control programs of web trojan and botnet generate the controlled end computer's programs,and successfully implanted in a controlled computer.The controlled Trojans need to find the corresponding control host in different ways,such as fixing IP address, dynamic domain name, IRC, P2P protocol, etc.In either way, like the controlled Trojan finding the source and the control panel sending the specified content,it must include source and destination IP addresses, ports, transport protocol, return values and other information,and the information can be found and recorded by means of monitoring.

3. The Triples Anomaly Detection Security Model

3.1 Safety Index of the Model

In this paper, we use the address entropy to reflect the distribution of the IP address of the attack event.The IP address is more confused and the distribution is more dispersed, the higher the address entropy;And the more ordered the IP address, the more concentrated the distribution, the lower the address entropy.In this paper, we will count the number of occurrences for the source IP address and the destination IP address.In order to count easily,the 32 bit IPv4 address is mapped to 16 bit integer by using Hash algorithm,so it can count the number of occurrences for the source IP address and the destination IP address with two arraies which number of elements is 65536.The calculation method of address entropy about the source IP address and the destination IP address is shown in the formula (1):

$$H = (-\sum_{i=0}^{65535}(\frac{C_i}{S})\log_2(\frac{C_i}{S}))/\log_2 S \tag{1}$$

In the formula:Ci--The number of IP address after operating by Hash;S--The total number of IP addresses in the current observation period, $S = \sum_{i=0}^{65535} C_i$.

Since all the benchmarks are subject to normal distribution, the actual value of the benchmark index should not much different compared with the statistical model in normal circumstance.Suppose for the benchmark index--S,parameters in the model are obtained by the model are $N(\mu_0, \delta_0)$,Where μ_0 and δ_0 are the mean and the variance of the normal distribution model.At present, the real-time value of this indicator is S,The method of judging the anomaly detection is shown in the formula (2):

$$\begin{cases} s-\mu_0 < 2.33\delta_0: \text{ Normal} \\ 2.33\delta_0 \le s-\mu_0 < 3.1\delta_0: \text{ Mild abnormal} \\ 3.1\delta_0 \le s-\mu_0 < 3.72\delta_0: \text{ Moderate abnormal} \\ s-\mu_0 \ge 3.72\delta_0: \text{ Severe abnormal} \end{cases} \tag{2}$$

For address entropy index,it should judged by $|s-\mu_0|$,because the address entropy value is too large (IP address distribution is too scattered) or too small (IP address distribution is too concentrated) are both exceptional.

3.2 Judging Events' Type by Triples

The rank about the statistical number N of the triples anomaly detection security model is 10-50.

The calculative content of the triples anomaly detection security model is following:

① Frequent events (the most frequently used attack event): the event set which events' type are same.

② Endangered target (the most attacked host): the event set which events' destination IP address are same.

③ The main attack source (the most active host from the external use of Trojan and other threat): the event set which events' source IP address are same.

④ Suffering from a kind of attack (the same target has been attacked the same way by multiple attackers): the event set which events' type and events' destination IP address are both same.

⑤ Varieties of attack methods (the attacker to try a variety of methods, repeatedly attacks against the same target): the event set which events' destination IP address and events' source IP address are both same.

⑥ The same attacking way (the attacker uses the same method to attack multiple targets): the event set which events' type and events' source IP address are both same.

⑦ Single mode attack (the attacker uses the same method, repeatedly attacks on the same target): the event set which events' type ,events' destination IP address and events' source IP address are all same.

3.3 Decision Tree Algorithm

In this paper, the decision tree algorithm is used to select a large number of data from the data.The decision tree is a prediction model showing the influence of the training data on the variables like a dendritic tree.According to the different of the utility of the target variable, the rules of the classification are constructed.

Decision tree construction is carried out in two steps.First step, generation of the decision tree:process of generating decision tree by training sample set.In normal circumstances,training sample data set is a date set based onthe actual needs of a historical and setting for data analysis at a certain degree of integration.Second step, decision tree's optimization:the optimization of decision tree is a process of testing, calibration and modification of the decision tree.The decision tree needs or not to be optimized according to the purity of the node classification.There are many ways to gauge the purity,the method used in this paper is judged by entropy.The method of judging entropy as shown in (3):

$$E = -\sum_{i=1}^{n} P(i) * \log_2 P(i) \tag{3}$$

In the formula: $P(i) = \text{Total number of class i} / \text{Total number}$

The smaller the entropy, the segmentation point is more accurate and the decision tree is more pure.When the entropy value is less than the specified threshold,the decision tree can be optimized by the way that generally choosen of triming branches.Decision tree algorithm is shown in figure 2:

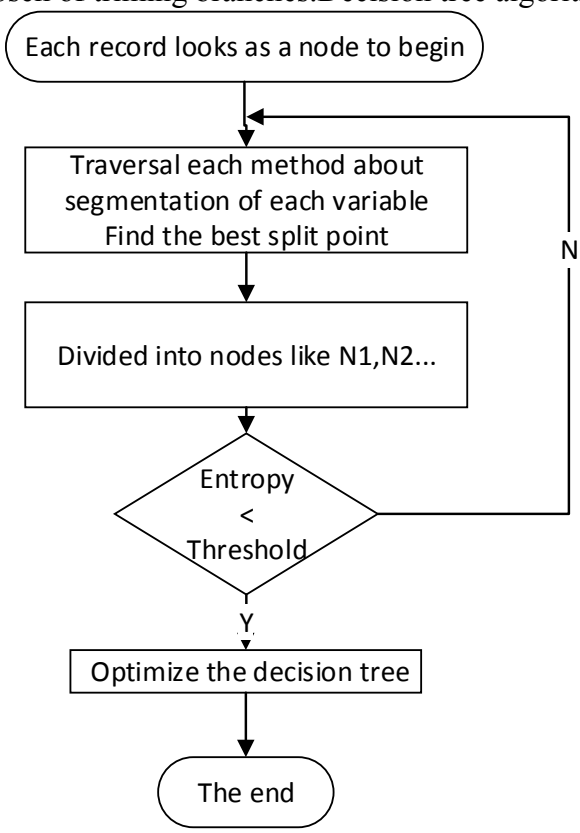

Fig. 2 Decision tree algorithm

3.4 The Flow of the Triples Anomaly Detection Security Model

For security threat events,data analysis of the triples is considered from the source, destination IP address and the types of threats in order to get the relevant conclusions with the most simple and efficient way.

The principle of the triples anomaly detection security model is collecting all kinds of data in network security equipment and filtering the data.After being calculated though the model, the data information is judged whether the event abnormal and the degree of abnormality.Then write these into the database according to the degree of abnormality.For severely abnormal events, a warning may be given and a record can be written,for preventing to influence the network security in the future.The frame of the triples anomaly detection security model is shown as figure 3:

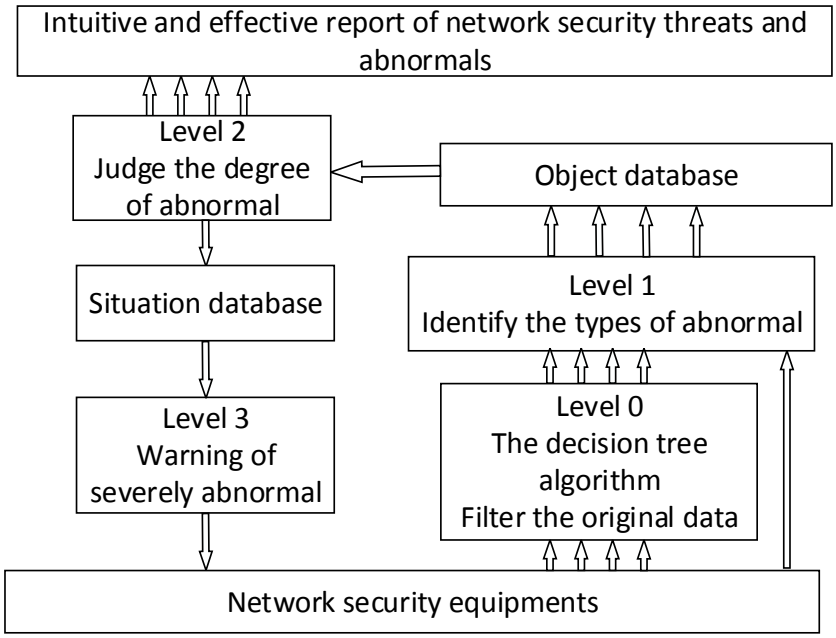

Fig. 3 Frame of the triples anomaly detection security model

According to analysis of the model frame,anomaly detection process of the model is following:

① According to the security event data provided by the network equipment, filtering a large number of data with decision tree algorithm to obtain effective parameters that model needs.

② The data which has been filtered by decision tree will be identified, classified and recorded into the object database according to different types of abnormal.

③ According to the requirement of model parameters,after analyzing the parameters of data sets from the database,determining its degree of abnormality and then writing the analysis results into the database and the reports.

④ It can remind tester the severe abnormal if necessary.

4. Examples of theTriples Anomaly Detection Security Model

4.1 Example of the Endangered Target Events

Figure 4 shows the change of the destination IP address though a period of time when there are three different events.

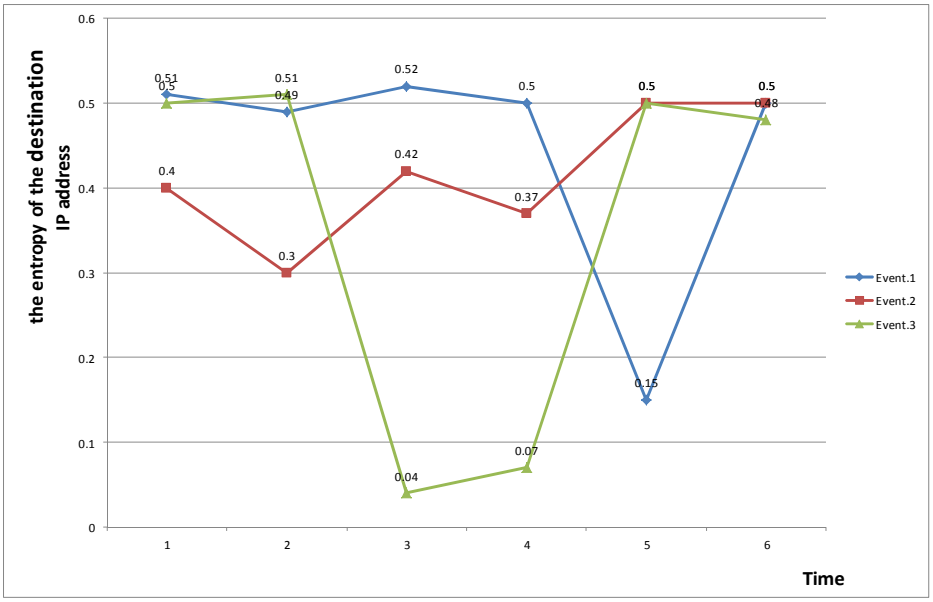

Fig. 4 Trend chart of the endangered target events

According to Figure 4, in case of that S was 0.5(S=0.5).It can calculate the following results: μ_1=0.445, δ_1=0.02099; μ_2=0.415, δ_2=0.00599; μ_3=0.35, δ_3=0.0524。

The following conclusions can be drawn:

① $(2.33\delta_1 = 0.0489)$<($|s-\mu_1|$=0.055)<$(3.1\delta_1 = 0.0651)$.There is a sudden drop of Event 1 at 5. Event 1's degree is mild abnormal.

② (($|s-\mu_2|$=0.085)>$(3.72\delta_2 = 0.0223)$.The entropy of Event 2's destination IP address has been on the rise. Event 2's degree is severe abnormal and must be taken more attention.

③ $(2.33\delta_3 = 0.1221) < (|s-\mu_3| = 0.15) < (3.1\delta = 0.162)$. There is a sudden drop of Event 3 between 3 to 4. Event 3's degree is also mild abnormal.

4.2 Example of the Frequent Events

Figure 5 shows the change of the source IP address though a period of time when there are three different events .

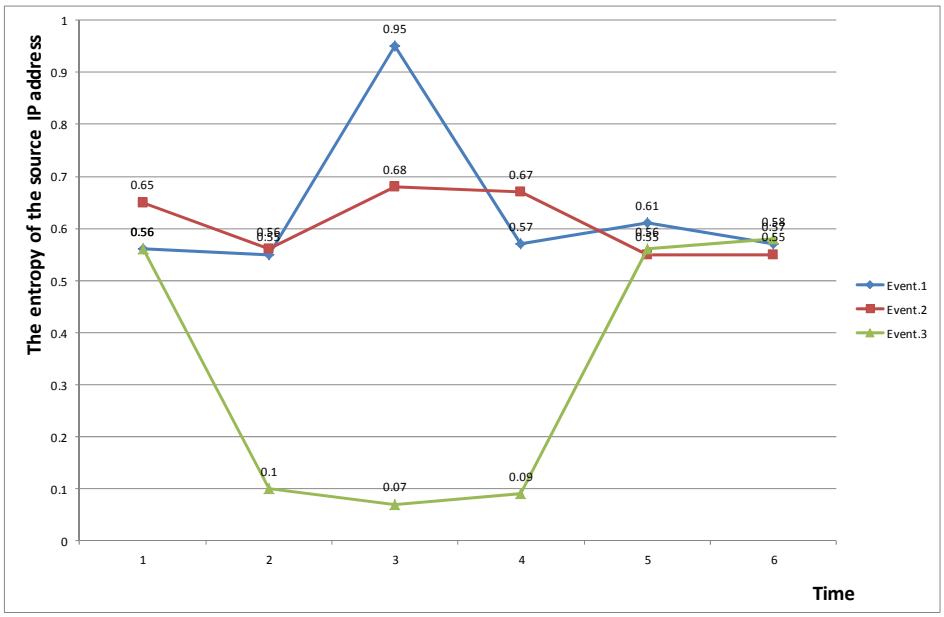

Fig. 5 Trend chart of the frequent events

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

According to Figure 5, in case of that S was 0.6(S=0.6).It can calculate the following results: μ_1=0.635, δ_1=0.02423; μ_2=0.61, δ_2=0.00396; μ_3=0.327, δ_3=0.06927。

The following conclusions can be drawn:

① $(2.33\delta_1 = 0.0567)$>($|s-\mu_1|$=0.035).There is a sudden rise of Event 1 at 3. Event 1's degree is normal.

② $(2.33\delta_2 = 0.00923)$<($|s-\mu_2|$=0.01)<$(3.1\delta_2 = 0.0123)$.The entropy of Event 2's source IP address has been in the stage of fluctuations. Event 2's degree is mild abnormal.

③ $(3.72\mu_3 = 0.2577) < (|s-\mu_3| = 0.273)$. There is a sudden drop of Event 3 between 2 to 4. Event 3's degree is severe abnormal and must be taken more attention.

5. Summary

This paper established a triple anomaly detection model for analyzing the source IP address, destination IP address and event type of network anomalies,and dividing the degree of abnormal level of security incidents.The results show,judging those events more rationally,classify the degree of abnormal more correctly,and give a warning or take more attention to severe abnormals with the triples anomaly detection security model based on decision tree.Parameters in the model only

three,but in future, it is neccessary for further digding out other available parameters for being composed a new data set to make model more comprehensive and practical.

Acknowledgment

This paper is funded by the project of The State Grid Corporation of China in 2014 “Research and development on the information safety threat analysis technology during the network access process”.

References

[1]. Yao QingFeng.Analysis of Network Security Situation based on Data Mining.Shanghai Jiao To ng University.

[2]. Ma Jie.Situation Assessment of Network Security Threats and Study of Analysis Method.Huah zong University of Science and Technology.

[3]. Lei Jie.Network Security Threats and Study of Situation Assessmen.Huahzong University of Sc ience and Technology.

[4]. Wang KaiZhuo.Research on the technology of network security threat situation assessment and analysis.Harbin Engineering University.

[5]. Chen RongMao.Research on threat modeling and detection technology of complex network.Na tional Defense Science and Technology University.

[6]. He QingBi,Hu YongJiu.Survey on Data Mining Technology.Journal of Southwest University f or Nationalities (Natural Science Edition),2003.06.30.

[7]. Zhang XueSong,Mao YunLong,Tan ZhuNan. Survey on Data Mining Technology. China Petroleum and Chemical Industry.

[8]. Liu JunQian.Research on massive data mining technology.Zhejiang University.

[9]. Cai Yan.Application of data mining technology in network information security management.N etwork Security Technology & Application.

[10]. Yang ZhengYi,Shen Hui,Li Xin.Information security thinking of big data.Electronics World.

[11]. BIBAK.JIntegrity Considerations for Secure Computer Systems[R].Bedford: ESD-TR-76-7 32, 1977.

[12]. Sun Mo. Research and design of P2P network security model. Xi'an Electronic and Science University.