The Honeynet Analysis and Assessment of Hierarchical Fusion Model based on Information Fusion

Xiujuan Yan [1, a], Qing Zhu [2, b]

[1]Xi'an University of Science and Technology, Xian, Shaanxi 710054, China

[2]The Active Network TM, Xian, Shaanxi 710068, China

[a]654676774@qq.com,[b]57481664@qq.com

Abstract. Using information fusion unique three layer structure, proposes a method of Honeynet analysis and assessment of hierarchical fusion model based on information fusion technology. The model uses the active defense technology of Honeynet real time tracking hacker intrusion behavior, and introduces multiple sensor information fusion thought; using intelligent algorithm automatic identifies a variety of network attacks finally in each layer. Firstly, analyzes the problems existing in the network security model, Then discusses the fusion model of levels of processing functions, introduces the technology, algorithm and problem resolved in the model.

Keywords: Information fusion, Honeynet, Network security.

1. Introduction

Network security has become more and more important and has become the extremely urgent problem to solve the Internet and many web services applications. For a network system, the goal of network security is to ensure the security of network data and work resources. The traditional network security model has many problems, detailed below:

1.1The existing problems in the network security model

Network security model generally includes data collection and induction, behavior analysis and classification, report and response and so on [1]. The current network security model has the following drawbacks:

First of all, the traditional information security technology mainly include intrusion detection system (IDS), firewall, virus protection technology, data encryption and authentication technology and so on[2], these techniques are important but there are limitations, they are completely defensive, in this kind of purely defensive strategy, the enemy has the initiative. Because as long as there is enough time, an intruder may detect provide service for the outside world from firewall, once the firewall is penetrated by attacker, these security tools are unable to provide further protection for the internet. For example, IDS attack information is provided only when it is attacked, but it is difficult that to get enough time to protect all the system which is easy to be attacked, and general IDS is hard to judge and predict the new attacks.

Secondly, the existing network security model exist real-time faults, the general model usually record their intrusion behavior after the hacking has done, but experienced hackers will often modify the relevant system log immediately after network intrusion, don't leave any trace, so it's hard to do comprehensive tracking of hacking. Therefore, network security situation can’t change along with the time and network’s change of internal and external environment, the security policy cannot be updated constantly.

Thirdly, the existing network security model mostly use data from a single source. Single of system structure and data source would cause the lacking of a comprehensive understanding of the whole security situation, it will lack of detection of complex attacks means of detection and analysis of the intrusion behavior is not accurate enough, lead to the high rate of false positives and non-response rates, so we need to find an effect algorithm to reduce the omission and the rate of false positives.

Finally, the existing network security model collect a large amount of data, the different data has different degree of misstatement, so it is easy to produce a large number of fault data, such as firewall log or IDS detection information [3], contains both attack information and the legal system activity information, therefore it must spend a lot of time and energy to screen the key information from thousands of alarm, in order to identify attacks.

In order to solve these problems, it needs to design and develop an effective network security model and select the appropriate algorithm, the model must be able to identify and track dynamic network activity in cyberspace automatically, and monitor all kinds of attacks in cyberspace.

2. Layered fusion model based on information fusion of honeynet analysis and evaluation

Throughout most of the existing problems in the network security model, we think the attention of existing network security system integrity is not enough, we should consider multi-sensor integration, use real-time data collection to find attacks, study the reasonable network framework to solve the existing network security model, this paper introduces the idea of information fusion, puts forward a honey-net analysis and evaluation layered fusion model.

2.1 The framework

Below, according to the data fusion framework put forward by White [4], this paper proposes a layered fusion model based on information fusion of honeynet analysis and evaluation, use different information fusion processing method in different stages of data abstraction, using its unique three layers of structure: data layer, feature, policy makers, make the system hierarchy, all kinds of parameter information organization are more clearly. The hierarchical model is shown in figure 2.1:

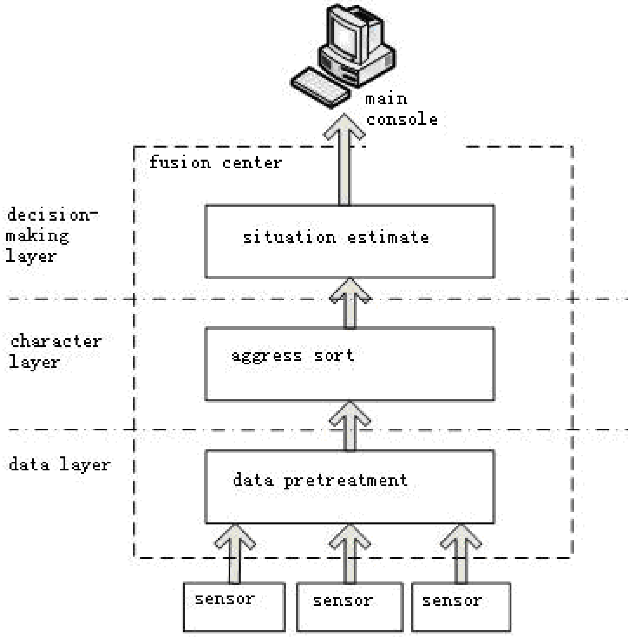

Fig 2.1 The Honeynet analysis and assessment of hierarchical fusion model based on information fusion

This model is a whole solution plan to analysis and evaluate the security of network, it use different information fusing processing methods according to different phases of data abstraction. Its main function is as follows:

(1)Sensor

Behavior data to capture the hacker is done by sensors in the system, sensor is the bottom and the most critical part of the whole system it is the basis of the data correlation analysis and hacking tracking. Data source cannot only rely on a layer to capture the information, it should collect data from various sources, so that it can reduce the risk of failure while increasing the information collected, the date source in honeynet model based on information fusion is the network nodes date collected by multi-sensor (such as security scanning tools, IDS, the honey-pot, firewall, VDS [5]), such as system log, warning, file, the audit data, network packets, etc. It can be represented as: St = {T1, T2…Tn}. In the formula, t is the time, St is the time information vector of t, Tj is the feature information of network node collected by the sensor. Sensors need capture data from the 4 layers: network, system, application activities and user activities.

The capture Standards listed below is based on honey-pot network capture data standard [6].

A. The captured network activity is collected by tcpdump binary format, and do a rotate/compressed (zip/ gzip) every day. Log is named year - month - day form.

B. Early warning of network intrusion monitoring use Snort 1.8 x full format.

C. We must do logging for each new started link, and archiving in (year/month/day) (military time) (source IP) (destination IP) (protocol) (source port) (destination port) (icmp_type) (icmp code) format. And all the port must be a numeric value.

D. The key in various attacks of the trend analysis and monitoring network is that each new different connection must be differently logged in a single source in the period of 24 hours. Different connection is defined as a kind of different source IP address. All port must be a numeric value.(year/month/day) (military time) (source IP) (destination IP) (protocol) (icmp_type) (icmp_code).

(2)Data preprocessing

This function is to do correlation and filter that suspicious event reported by each sensor, encapsulated into the next layer analysis after parsing and unified format, to facilitate the subsequent work. This layer corresponding to primary filter layer in the White model, different sensors capture the different information format, using the data pretreatment Agent to handle the information collected by the sensor for each sensor, the Agent use an XML encapsulation and SSL encryption transmission for data processing, translate behavior data newly produced in the log into standard behavior record through real-time parsed, finally send this information into the fusion database.

For system log files, data collection Agent is used to collect a host of audit records, system log and other data, data pretreatment Agent is used to simplify the log file collected, submit to fusion database according to the unified format; Network data collection Agent mainly collect features associated with the network connection, such as protocol type (protocol type), dst bytes (destination address to the number of bytes of source address), src bytes (source address to the target address number of bytes), service (target web services), etc. these features can reflect the type of the connection, and improve the efficiency of network intrusion detection. Each time after data collection Agent intercept a packet, unpacking, and then send to pretreatment Agent. Data preprocessing module structure is shown in figure 2.2.

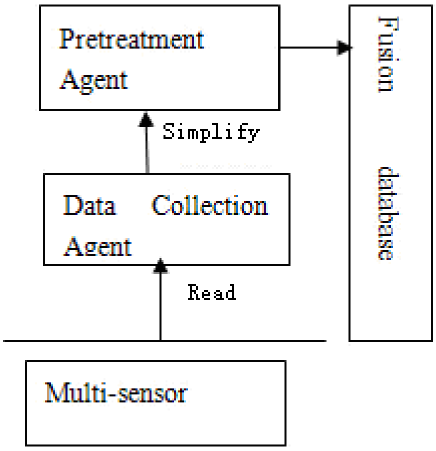

Fig 2.2 data pretreatment module

(3)Attack classification

The module corresponds to the primary layer treatment of White model, its function is the fusion of suspicious event reported by multiple sensors, reduce the false positives through multi-sensor integration. Use data mining technology to processing cluster data, module structure is shown in figure 2.3.

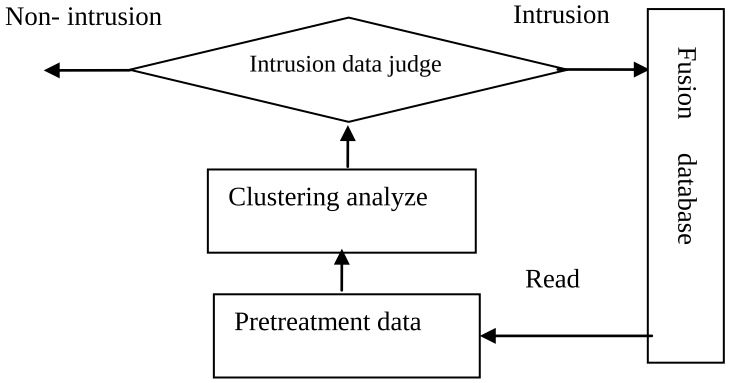

Fig 2.3 attack classification deal with module structure

The core of this module is to find attack behavior rules in a large amount of data, data mining technology is the first choice of this module, and there are great advantage in extracting features and rules from the data. As the object of the honey-net study is a large-scale network, involving multiple snort, honey-pot distributed in each segment, , its information including source of attack, goal of attack, attack types, attack times, and other multiple attributes, so the multi-source information fusion technology is a better choice.

In information fusion method, fuzzy clustering technique is a good way to solve the problem of classifying ambiguity, the uncertainty and fuzziness question that there are no clear boundaries between class and class of uncertainty. As a result, according to clustering method, the module divide data set into the invasion class and normal class sets, we only consider invasion, and this part of data will fusion into the next link that is policy maker layer.

(4)Situation assessment

This module corresponds to the situation assessment and threat estimate layer of White model. Its function is to get the current system analysis of the invasion based on the lower attack report and comprehensive historical situation information in the database [7]. Situational assessment is to evaluate the safety status of existing systems and development trend forecast. Situation assessment is a kind of reasoning of policy makers, it need to be accurate, completely percept each other's intentions and tell each other’s attack plan, so as to provide decision support for security personnel. Situation assessment needs to forecast and analysis calculation with the information change constantly, so it is a dynamic process, that is known the situation A (t) of t moment, to get network safe operation of A Δ (t + t) moment. Situation assessment process is to create maps the process of phi. Φ: A (t) - AΔ (t + t) [8].

Due to the uncertainty and incompleteness of knowledge and information, situational analysis process need to rely on rich domain knowledge, therefore the network security situation assessment is under the hypothetical inference framework, according to incomplete, uncertain or imprecise of the knowledge and information make reasoning decision-making.

The object of threat estimation research in this model is to predict attack collection, each attack include the attack source and target, attack port, attack times and other properties, every attack set consists of one or more such attacks, so we should apply the algorithm of combining attribute weights and fuzzy evidence theory, it can well solve the problem of multiple attack type identification inconsistent, fused the recognition results effectively, so as to get higher recognition correct rate.

2.2 The application of Honeynet technology in the model

Through studying the problems of the existing network security model, we proposed a layered model Honeynet analysis and evaluation based on information fusion, the current study is only to combine the intrusion detection systems and information fusion. In order to change from the current passive defense to active defense system; The current information collection methods converse only through a single IDS alarm to multiple sources obtain information, our applied Honeynet system in the model, so that more accurate and effective capture data analysis [9]. Here we explain what a Honeynet is.

Honeypot is a system that used to observe how hackers detect and eventually invade the system, load some dates and applications in the system that does not threaten the company. But for the hackers it is a great temptation, for example, a computer in the network, it seems that the general appearance of a single machine, but through some special configuration to attract potential hackers and capture their traces, this computer is like a mousetrap.

Honeynet belongs to high interaction of honey-pot, which attract the computer network that is attacked by the hackers. It generally consists of firewall, router, intrusion detection system (IDS) and one or more network systems. Honeynet technology has changed the traditional model of passive defense network security situation, it pretends to a set of network servers among the normal network to lure hackers, to gather information which is threaten safety, to find the attacker's motivation and determinate the attack mode, that is, find the new intrusion tools. It is used to gather information to better deal with threats from within or outside.

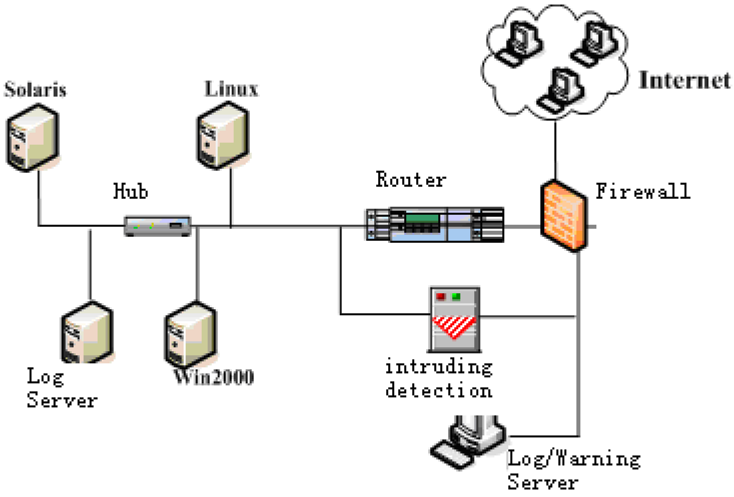

Fig2. 4 Honeynet structure

Honeynet can track invasion of hackers in a real-time, figure 2.4 shows the detailed structure of honeynet. The figure includes three different networks: the honeynet, management networks and Internet. Win2000, Linux, Log server is for the Honeynet; log/warning server is for network management. All the packet is through the honeynet must go through routing and firewall. Firewall main control in and out of the connection, it is designed to allow all incoming connections, but control the output connections. It can track the connection which initiates the Internet from the Honeynet. If honeynet reached the number of connections to send out, all subsequent connections attempt will be blocked by the firewall. This provides the opportunity to the hackers to do what they want to, while being able to prevent the system from being abused. Routers are used to supplement the firewall filtering [10].

Because there is no authorization service of honeynet, so any interaction with the honeynet are malicious or non-authorized, any connections from honeynet indicate that someone break into the system and initiate activities outside, any connections into the honeynet belong to scan, probe or attack, any information from honeynet captured that related to the attack, this makes it easier to analysis the honeynet activity inside.

In summary, we combine honeynet with information fusion technology [11], can be more efficient access to network data. We introduced the honeynet system, you do not just rely on IDS alarm to get information and through multi-sensor collection of hacking real-time tracking, and you can get more valuable information [12].

2.3 The composition of honeynet based on information fusion

In order to build a security model, it should include firewall, honeypot, intrusion detection and other security components in the system, provide for the network to prevent, detect attacks and response capabilities through their joint work. And the appropriate algorithm should be selected, detect the data collected in the network, help network security personnel analyze and take preventive measures. Honeynet system based on information fusion is proposed in this paper, using information fusion model, selecting estimation and identification fusion technology to strengthen greatly the information analysis function of the honeynet, it will be able to provide more in-depth safety information [13]. The system includes a component shown in Fig 2.5.

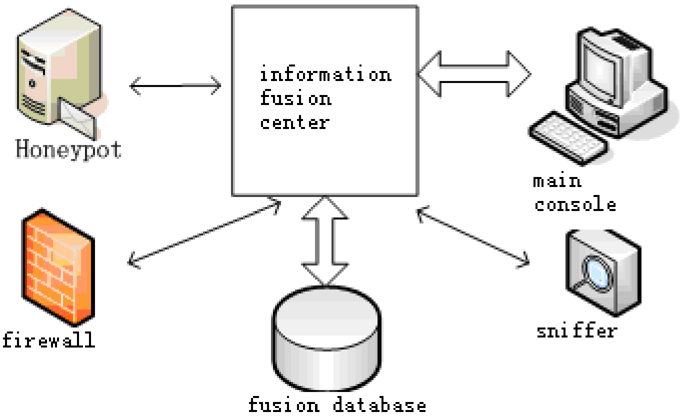

Fig 2.5 the composition of Honeynet based on information fusion

(1) The function of honeypot in the honeynet based on the information fusion

Honeypot can lure hackers, gather information on security threats, it can detect and respond to proactive network of intrusion and attacks. The honeypot is not just one but multiple in honey nets, they distribute in different segments. Honeypot system is a real system, no simulation environment or deliberately.

(2) The function of firewall in honeynet based on the information fusion

As a threshold, firewall limit's internal and external network to access to each other, its configuration of filtering rules to prevent unauthorized users access to the internal network. So when the system detects attacks and automatically modifies the firewall security policy, this can block suspicious network communication. In addition, since all traffic must flow through from it, and the hacker does not detect, the firewall has a good ability to capture the data. Firewall logs mainly records header information, such as date or time of attack, source and destination IP addresses and source and destination ports.

(3) The function of IDS in honeynet based on the information fusion

Intrusion Detection System (IDS) has two important functions, one can capture all network activity, which can capture and record every packet and load, but to capture so much information will consume a lot of resources, most organizations cannot afford it [14].However, honeynet can easily capture all packets activities, because of honey net's few activities, it often does some forms of attack or detection, so honeynet can analysis attacks at the network layer. And it can capture saddlebag, keystroke behavior, even the communication between the hackers. The captured information is usually preserved by IDS in binary files.

The second function of IDS is to alert for any suspicious behavior. But because all inbound connections are regarded default as possible or malicious, for honeynet, IDS alarm has little value and not critical.

2.4 The problem solved by the model

By analyzing the existing problems of network security model, we introduce ideas of information fusion, we proposed a honeynet analysis and evaluate model, using intelligent fusion algorithm to analyze the data of network security situation in each layer. The system uses active defense technology, are more easily to get the key data we needed, it can reduce the false alarm rate, to solve the data acquisition unity, improve data collection comprehensive, solved the problems such as the non-real-time, and can assess accurately threat, etc. [15].

First, use the active defense technology to reduce the false alarm rate.

The Honeynet technology, the purpose is to deliberately tempt the hacker's attack, which in the case of not being noticed monitor hacker's action, in order to understand and study the hacker's motivation, methods and technology adopted. In addition, most of the monitoring methods will produce a lot of false data, such as intrusion detection system, the validity of this model can be greatly reduced, and with the honeynet captured data with very low misstatements almost all incoming traffic honeynet system reflects a type of malicious activity, the analysis of such data reliability can be greatly increased [16].

Second, solve the data acquisition of singularity.

Using multi-source sensor technology can collect and capture more data, the equipment which is used to capture data is usually the firewall, intrusion detection system and honeypot host (honeypot). Data capture is intended primarily to capture firewall logs, IDS logs and honeypot host system log, or "triple capture." Capture as much information attacks, and fully guarantee the integrity and security of information capture.

Third, solve the non-real-time.

Adopt the method of real-time data processing, able to track from the connection of the system to leave all hosts in the whole process of operation behavior and the network connection of the hackers, and on this basis, to research these data and get the real-time trend analysis of data.

Fourth, assess the threat accurately.

This model uses information fusion technology, adopting layered fusion method avoids the uncertainty of a single analytical method, layered fusion will be hierarchical intelligent information analysis, feature layer using clustering algorithm to classification the attack, in the policy makers through fuzzy D - S evidence combination to fusion of data analysis results, and rough set theory is introduced to compute the weight of each attribute, and ultimately determine the trend of network security at time T.

3. Conclusion

First of all, study the problems existing in the existing network security model, according to the present situation research put forward a honeynet based on the information fusion analysis and evaluation of layered fusion model. The model dynamic collection of information from multiple sensors, the information is of continuity in time and is of crossing in space, reduce the false alarm rate and improve the comprehensiveness of the data collection. Use the information fusion model for layered fusion to get accurate threat assessment by hierarchical intelligent analysis. This article made a detailed functional description to the relevant module model.

References

[1] He you, Wang guorong, Peng yingning. Multi sensor information fusion and application, Beijing: Publishing House of electronics industry, 2000.

[2] Liu tongming, Xia zuxun, Xie hongcheng, Data fusion technology and its application, Beijing: National Defense Industry Press, 1998.9

[3]Zhang Yu, Research and design of honeypot system, Computer engineering and Application, 2006.10

[4] Lawrence A. Klein, Sensor and Data Fusion Concepts and Applications, 2004.

[5] Liu wei, Liu lu, estimate the safety situation of fuzzy pattern recognition based on D-S evidence theory, Computer engineering and Application, 2006.

[6] Zheng xiaoyong, Yao jingshun, Fuzzy pattern recognition method based on D-S inference, Systems engineering and electronics, 2003.

[7] Glenn Shafer. A mathematical theory of evidence [M]. Princeton University Press, 1976

[8] CHEN SHOU-YU. Fuzzy recognition theoretical model [J]. The Journal of Fuzzy Mathematics, 1993, 1(2) p.261-269

[9] Li mingjun, Liu yixin, Battlefield target identification based on fuzzy pattern recognition, Fire control and command, 2005.12

[10] Liu weiyi, Liweihua, Intelligent data analysis, Beijing; Science Press, 2007.9

[11] Liu Mixia, Zhang Qiuyu, Network Zhao Hong, Yu Dongmei, Security Situation Assessment Based on Data Fusion, Knowledge Discovery and Data Mining, 2008. WKDD 2008. International Workshop on Volume, Issue, 23-24 Jan. 2008 Page(s) p.542 – 545

[12] WaldL. Some terms of reference indata fusion geoscience and remote sensing [J].IEEE Transactions on Geoscience and Remote Sensing, 1999, 37(3) p.1190~1193.

[13] Ge haihui. Lu xiao. Zhou zhenyu. Data fusion technology in the network security management platform. Modern electronic technology, 2004. 24 p.69~70

[14] Wierman M J. Measuring Conflict Evidence Theory [A]. IFSA World Congress 20th NAFIPS International Conference[C].2001.p.1741~1745.

[15]Wang zuli. Gan gang. Data fusion technology apply in the network security. Fujian computer, 2007, 5:p.79~80

[16] White F.A model for data fusion[C] SPIE Conference on Sensor Fusion Orlando, FL. Aprial, 1998.