A Multi-source Alarm Information Fusion Processing Method for Network Attack Situation

Jiawei Du1,a, Ronghua Guo1, Guowei Suo1 and Xing Zhang1

1LEETC, Zhoushan Road 17#, Luoyang City, Henan Province, China

a Jiawei Du: 15236132569@163.com

Abstract. Data fusion technology is one of the key supporting technologies for network security situational awareness. This paper focuses on the research of multi-source alarm information fusion processing method, analyzes the level of network security situation perception data processing, and gives the process of processing data using data fusion technology. It takes the alarm information of network security equipment as the data source, and puts forward the theory of attribute similarity clustering and weighted D-S evidence theory. Source data fusion and vulnerability information association analysis are three ways to integrate multi-source alarm information fusion for network attack situation. This method takes a comprehensive consideration of the alarm information with a number of devices, and filters and fuses the alarm. It can reduce the number of alarm and make the final warning results better reflect the attack situation in the network.

1. Introduction

Attack situation is an important aspect of network security situation awareness. There are a lot of attacks in the network. An attack often triggers alarm on multiple security devices, such as IDS, firewall and anti-virus software. The fusion processing method is used to fuse the alarm information of different devices, and effectively make use of the security information provided by these security devices to complement each other, which can reduce the redundancy of the alarm. At the same time, many network attacks are based on vulnerability, and the attack information is associated with vulnerability information. It can further judge the credibility of the attack, and make the detection of attack information more accurate.

In the network, the alarm information caused by the attack is very large. In view of the data provided by various security sensors, a good processing method is needed to deal with the alarm information synthetically so as to prevent the useful security information from being flooded in a large number of noises. A suitable and efficient processing method can reduce the processing of redundant alerts by security managers and improve the timeliness of information processing. In this paper, a multi-source alarm information fusion method for network attack situation is designed. Through clustering, fusion, association analysis and other means, the security alarm information of different data sources is processed synthetically, which makes the number of warning of the final report greatly reduced, and can reflect the attack situation in the network truly.

2. The Related Theory of Network Security Situation Awareness

2.1. Situational Awareness Data Processing Level

There are many reference models in the field of network situational awareness. Based on the needs of process data processing, this paper divides the corresponding data processing process into five
levels[1], which are data collection, data preprocessing, information extraction, situation analysis and situation presentation, as shown in Figure 1.

(1) Data acquisition refers to the acquisition of all kinds of data related to network space security from various devices, such as system log, IDS alarm, vulnerability information, network topology and so on.

(2) Data preprocessing is the initial processing of the acquired data. Because all kinds of data are derived from different devices and have different formats, the data are cleaned, integrated, reduced and transformed through this process, and the data are fused to multisource heterogeneous data.

(3) Information extraction is a further fusion understanding of data produced by data preprocessing. By establishing appropriate situation awareness index system, the data of different sources can be fused to generate the underlying index.

(4) Situation analysis is the integration of the underlying situation indicators, through a variety of data processing methods integrated processing and calculation, and then get the upper class situation results.

(5) Situation display is the display of cyberspace situation through suitable visualization means for users to make comprehensive judgments.

2.2. Process Flow of Situational Awareness Data Fusion

In the five levels of data processing in situational awareness, we can see the extensive and frequent applications of data fusion technology, which are involved in (2), (3) and (4). In the two levels of data preprocessing and information extraction, different data fusion algorithms are used to deal with multi source data, extracting and fusion from the data, and getting the corresponding situation indicators[2]. In situation analysis, a large number of data fusion techniques are also useful. Its corresponding data processing flow framework is shown in Figure 2.

Figure 1. The Level of Data Processing in Situational Awareness

Figure 2. The Data Fusion Process of Situational Awareness
1. **Data preprocessing**: different data sources are different for the same target in the network space, such as various vulnerability scanning tools, IDS devices and other data formats are different. Aiming at the security situation index in cyberspace, the heterogeneous and heterogeneous description information is obtained. To make full use of these information, first of all, the data should be calibrated and normalized, and the corresponding information should be unified according to the situation indicators. For example, for alarm class data, cluster analysis can be used to realize alarm clustering and simplify alarm.

2. **Information extraction**: further fusion processing of the output data after the data preprocessing is carried out to get more accurate and comprehensive situation index data for situation analysis. At the level of information extraction, it is a further fusion of data processed by the data processing layer, and the fusion algorithm of response is used to fuse the processing index data so that the obtained data are more comprehensive and accurate. For example, we can integrate multiple features through D-S evidence theory to comprehensively evaluate network traffic, and achieve more accurate judgement of abnormal traffic.

3. **Situation analysis**: the data of information extraction and output are processed again, and the fusion algorithm and evaluation calculation method are used to process and calculate all kinds of low level situation index data, and then the situation results of the upper level are obtained, which can be used as the basis to judge the overall network security situation.

   It can be seen that data fusion is the basis and core of situational awareness. How to integrate multi-source data effectively is directly related to the perception of performance and accuracy. In this paper, the fusion algorithm is used to fuse the multi-layer and multi source data in the network space, and the complex alarm data will be processed comprehensively to reduce its complexity and redundancy as the basis of situation analysis and situation presentation.

3. **Multi-Source Alarm Information Fusion Processing Method**

   In this section, a method of processing alarm information by clustering, fusion and association analysis is designed to get an alarm that can truly reflect the threat of the network. The process is shown in Figure 3.

   ![Figure 3. The Fusion Process of Alarming Information.](image)

   (1) Alarm clustering is to register and cluster the alarm information of different devices. A large number of the same or similar alerts produced by different security devices are merged into a super alert, thus realizing the simplification of the alarm. In this paper, alarm clustering algorithm based on attribute similarity is applied to alarm clustering[3].

   (2) The alarm fusion is to fuse the alarm of different devices by information fusion algorithm, to further confirm the alarm by the complementarity of the information between different security devices, so as to more accurately locate the intrusion and reduce the number of alarm. In this paper, we use weighted D-S evidence theory to fuse different device alarm information.
(3) Association analysis is the correlation analysis of the integrated alarm information and the vulnerability information existing in the environment. The alarm is further screened by associating vulnerability information to remove invalid alarm. In this paper, we analyze the vulnerability of the alert and the vulnerability list of the attack target, so as to determine whether the attack of the representative is successful.

3.1. Alarm Clustering Based on Attribute Similarity

In actual attack detection, one attack will produce multiple alarms on multiple devices, especially IDS and firewall, and often produce a lot of the same alarm in a short time. To fuse the network attack data, the alarm data should be processed first, the alarm of the equipment is clustered, the repeated alarm is merged, and the super alarm (hyper-alert) is formed as the processing object of the subsequent fusion.

The alarm triggered by the same attack usually has the attribute similarity, so this paper uses the similarity based alarm clustering algorithm, which is based on the similarity of the warning attributes (such as IP address, alarm time, alarm plug-in number, alarm type, etc.), and extracts useful attributes from the alarm information. Then calculate their similarity and fuse enough similar alerts into one alarm, so as to reduce the number of repeated alerts and simplify the number of alarm.

As the original alarm information contains many attributes that are not related to the alarm clustering, such as the type of protocol and the result of verification, it is necessary to extract the alarm information first and extract useful attribute values, and then cluster it on this basis. According to the alarm of IDS, we can use the attributes of alarm time, source IP address, destination IP address, source port, destination port, alarm type, alarm plug-in number and other attributes to cluster. In view of the alarm of the firewall, we can cluster the alarm time, the source IP address, the destination IP address, the source port, the port and other properties [4].

The idea of alarm clustering algorithm based on attribute similarity is as follows:

(1) The time attribute is not suitable to be the component of the similarity calculation with other attributes, but it is suitable as a separate threshold to set the range of alarm clustering first, and the alarm within the time threshold is used to cluster the rest of the attributes.

(2) In the alarm clustering process, a comparison queue is set up for the extra alarm data generated for temporary storage. Each record in the queue has a field Count to store the same number of similar repeated alerts for the integration of the record, a Starttime from the beginning of the record alarm, and an Endtime field to record the deadline of the alarm.

(3) When the new alarm comes, the weights of each attribute and the comparison threshold of time attribute are set according to the alarm category. First, the time similarity between the time stamp of the new alarm and the Starttime of the superalarm is calculated. If the time similarity is 1, then the super alarm is reported and no longer participates in the subsequent alarm clustering process. The distance of the time attribute is 0, and the similarity between the alarm and the super alarm is calculated and recorded.

(4) To compare the alarm and all the super police in the queue, if the minimum similarity between the alarm and the super alarm is less than the threshold, the alarm is merged into the super alarm with the smallest similarity, and the Count value of the corresponding super alarm is added to 1, and the time stamp of the latest aggregated alarm is updated. Otherwise, the alarm will be added to the queue as a new super alarm. The Count value is set to 1, and Starttime and Endtime are set as the timestamps of the alarm.

The algorithm determines whether the new alarm belongs to the same alarm class, and the judgement depends on the set time threshold and the similarity of the alerts. The weights of different kinds of attacks can be different. We can adjust the weight of the warning attribute more reasonably by the method of experimental learning, and set the appropriate weight to make the algorithm more ideal for the clustering effect of different kinds of attacks.

3.2. Alarm Fusion Based on Weighted D-S Evidence Theory

After the clustering of different security devices, we will merge the alarm through the information fusion algorithm and confirm the alarm by the complementarity of the information between different security devices, thus further locating the network intrusion attack.
As the original D-S evidence theory is integrated, the evidence provided by different sources is equal to trust. In the actual environment, the credibility of different security devices is different. Therefore, this paper uses the weighted D-S evidence theory to merge the alarm. In the fusion process, the weight is used to indicate the trust degree of the corresponding equipment, which makes the fusion result reflect the real situation better.

The concrete steps are as follows: first, the basic probability distribution of different evidence is averaged, and then the information is fused by Dempster combination rules, and the degree of evidence supported by other evidence is used as the weight value. On this basis, the evidence is weighted averaging. Finally, the combination of Dempster rules is used. In the actual application environment, the reliability of different types of attack alerts can be obtained according to the experience of the experts or related knowledge, which can be used as the weight value for the fusion calculation. Compared with the classical D-S evidence theory, the algorithm divides the Dempster synthesis rule into two parts[5].

1. Distribute the basic probability distribution of N different evidence according to the weight average.

\[
m(A) = \sum_{i=1}^{N} r_i m_i(A)
\]

2. After calculating the M (A), the N \( m(A) \) is fused by using the Dempster synthesis rule.

The super alarm after clustering detected by the security device is used as the fusion object, and the recognition framework is defined as \{true, false\}, in which the true indicates that the alarm correctly reflects the attack, and the false indicates that the alarm is a false alarm, and the attack reflected is not happening. The basic probability distribution of alarm for N security devices is \( M_{1..N}(\{true\}) \), \( M_{1..N}(\{false\}) \), \( M_{1..N}(\{unknown\}) \), respectively, indicates the occurrence of A attacks, the occurrence of unknown attacks and the unknown reliability of N security devices. The weighted D-S evidence theory is used to fuse the basic probability distribution of the alarm of N security equipment, and the basic probability distribution values under the combination of multiple evidence are obtained. The probability distribution is used to judge the authenticity of the alarm. The corresponding decision rules are as follows:

\[
\begin{align*}
    & m(true) - m(false) > \varepsilon_1 \\
    & m(unknown) < \varepsilon_2 \\
    & m(true) > m(unknown)
\end{align*}
\]

The specific algorithms are as follows:

**Input:** Alarm clustering results for different devices to coincide with the same target time.

**Output:** An integrated alarm.

1. According to the relevant knowledge, we get the basic probability distribution of alarm device K by I.
2. Get the weight assignment of the security device I to the corresponding K attack.
3. Select the threshold for the fusion decision.
4. Initialization of the average probability distribution.
5. Calculate the average probability distribution of N devices.
6. \( M(\text{true}) \), \( M(\text{false}) \) and \( M(\text{unknown}) \) are the probability distributions after fusion, and are initialized to mean probability distribution.
7. Use Dempster fusion formula to carry out n-1 fusion and calculate the final result.
8. The judgement threshold is correct, and the alarm is correct.
9. Otherwise, the alarm is false alarm and will be filtered.

### 3.3. Vulnerability Information Correlation Analysis

In the actual application environment, the success of network attacks depends on the state of the target, such as whether the target is online, the vulnerability of the attack is on the target, and so on. Through
the correlation analysis of the corresponding target vulnerability information, the alarm data can be further filtered.

To link vulnerability, we need the following supporting information:

1. **Network attack dependent vulnerability information**

   Network attack depends on vulnerability information, which is a set of vulnerabilities needed for successful attack. Some attacks need to exploit the vulnerability of the target system or the vulnerability of the communication protocol, and some attacks do not require a vulnerability on the target, only the target is required to open the corresponding port and service. Taking IDS as an example, the description of the detection plug-in has the corresponding vulnerability information, which can ensure the correctness and completeness of the information by establishing the corresponding database[6].

2. **Target vulnerability information**

   We can scan the target system in many ways to get the vulnerability information of the target system. By integrating the scanning results of different scanning devices, we can get a more comprehensive and accurate list of vulnerabilities to reflect the target vulnerabilities.

   By matching the vulnerability and target vulnerability information of the network attack, the probability of the attack can be judged, and the further filtering of the attack information can be realized. For a specific network attack, the result of vulnerability matching can be calculated by the following formula:

   \[
   P = \sum_{i=1}^{n} R_i P_i
   \]

   Among them, \( n \) is the number of vulnerabilities that attack dependent vulnerabilities and target vulnerabilities exist. \( R_i \) is the corresponding weight of vulnerability \( I \), and \( P_i \) is the probability of existence of vulnerability \( I \).

   The result of the vulnerability matching is calculated and the \( P \) value is obtained. For alarm independent from loopholes, the \( P \) value is 1. For those without corresponding matching vulnerabilities, the \( P \) value is 0.1. Based on the \( M \) (true) and \( P \) values of attack, the probability of attack success is calculated: \( S \):

   \[
   S = M(\text{true}) \times P
   \]

   The threshold value of \( S \) is set and filtered, and the result is compared with the threshold value. If the result is greater than the threshold, it can be considered that the attack is likely to be successful and the alarm is output in time. If the result is less than the threshold, it is considered that the attack is unlikely to be successful and the attack information is filtered out. Through the analysis of the correlation with the vulnerability information, the output alarm can reflect the actual situation of the attack on the network, and the judgment obtained by the alarm is more accurate.

   In fact, the attack information that is not filtered by the vulnerability has certain reference value. They reflect the tendency of attack activities, together with the reported alerts, which reflect the utilization of vulnerabilities in the system. For those vulnerabilities that are often attacked, it is necessary to query whether other targets in the network have corresponding vulnerabilities and patch them in time [7]. The process is shown in Figure 4.
Alarm information of fusion output

A list of vulnerabilities to get the alarm target

Getting the vulnerability information of the alarm dependent

Calculate the matching probability $P$ of the vulnerability combined with $M$ (true) value, calculate the value of attack success probability $S$

Report a warning

Filtering alarm, recording vulnerability information can be used as a reference for patch.

Figure 4. The Process of Association between Alarm Information and Vulnerability Information.

4. Experimental Results

In order to verify the feasibility and effectiveness of the proposed method, we built a simulation environment specially for simulation experiments. First, a network topology is built, and then the ms10_046_shortcut_icon_dllloader attack module of Metasploit software is used to attack the host in the network, and the Snort and a certain type of firewall are used as security devices in the network to test the attack. The two steps of clustering and fusion are used to obtain the alarm information of the attack, and then the association analysis is made by using the list of target vulnerabilities to get the alarm information to be processed.

The ms10_046_shortcut_icon_dllloader module takes advantage of the CVE-2010-2568 vulnerability, through which local users or remote attackers can use a specially made Windows shortcut file (.LNK or.PIF) to execute a malicious DLL icon (.LNK or.PIF) to execute arbitrary code. The module set up a Web DAV service. Remote users trigger attacks when accessing specific URLs by IE, and can be used to execute arbitrary payloads.

The successful attack of this module requires that the target machine access specific URLs. This article simulates the attack process in the experimental environment. When the attack server accesses the target address, the address is redirected to the malicious DLL directory, which triggers the attack. The attack action triggered the alarm on the Snort and firewall, in which 22 alerts were triggered on the Snort, and 6 alerts were triggered on the firewall. Using this method, the results of corresponding alarm fusion processing are given, as shown in Table 1 and Table 2.

Table 1. The experimental data of clustering fusion results

| Sid   | Cluster amounts | $M$(true) | $M$(false) | $M$(unknown) |
|-------|-----------------|-----------|------------|--------------|
| 1:1042:16 | 18              | 0.78      | 0          | 0.21         |
| 1:17402:3 | 3               | 0.78      | 0          | 0.21         |
| 139:6:1    | 2               | 0.55      | 0          | 0.45         |
| 1:15330:7  | 1               | 0.55      | 0          | 0.45         |
| 122:1:1    | 1               | 0.32      | 0          | 0.68         |
| 1:1460:9   | 1               | 0.92      | 0          | 0.08         |

Table 2. The experimental data of vulnerability associated results

| Sid   | CVE           | Vulnerability associated results | Attack probability | threshold |
|-------|---------------|----------------------------------|--------------------|-----------|
| 1:1042:16 | CVE-2002-0756 | 0.2                              | 0.08               | Filtered  |
| 1:17402:3 | CVE-2010-2568 | 0.98                             | 0.72               | Reported  |
| 139:6:1    |                | 1                                | 0.55               | Filtered  |
| 1:15330:7  |                | 1                                | 0.56               | Filtered  |
| 122:1:1    |                | 1                                | 0.49               | Filtered  |
| 1:1460:9   |                | 1                                | 0.92               | Reported  |
After processing, the reported alerts are 17402 and 1460 respectively. The former "WEB-CLIENT Microsoft LNK shortcut download attempt" reflects the vulnerability of the attack, and the latter reflects the alarm triggered by the attack load.

In this paper, we use the real attack warning data to verify the effectiveness of the method by using the alarm triggered by the real attack in the network environment. The experimental results show that the alarm information can be reduced by the method proposed in this paper. At the same time, the alarm can accurately reflect the attack behavior in the environment, and provide a good data base for the estimation analysis of the attack situation.

5. Concluding
In this paper, we propose a processing flow of Clustering Fusion correlation vulnerability information to deal with the alarm information triggered by attacks in the network. By clustering, merging similar alarm, reducing the number of alerts that need to be processed; using the extended D-S evidence theory to fuse the alarm of different devices, effectively uses the complementary redundancy information of the equipment to filter the confirmation alarm; the reported alerts are associated with the vulnerability information of the target, and judge the attack according to the environmental information. The credibility of the work further filters out the unsuccessful attack alarm. Make the final report alert more accurately respond to attacks in the network. Experiments show that the method achieves the corresponding purpose, and the number of reported alerts is reduced, and the corresponding attack behaviors in the network can be truly reflected.

6. References
[1] Gutta S, Qi Cheng. Data-based Distributed Classification and Its Performance Analysis [C] 15th International Conference on Information Fusion, July 2012.
[2] Sushil J, L. Peng, S. Vipin, W Cliff. Cyber Situation Awareness-Issues and Research [M]. Springer New York Dordrecht Heidelberg London 2010.
[3] The MITRE Corporation About CVE[EB/OL]. http://cve.mitre.org/about/index.html.2008-06-04/2009-03-30.
[4] MiaoWei, Liu Yanhua. Information System Security Risk Assessment Based on Grey Relational Analysis and Dempster-Shafer Theory [C]. 2011 International Conference on Mechatronic Science, Electric Engineering and Computer. August, 2011.
[5] Wang Yan network security event aggregation and association analysis technology [D] Harbin: Harbin Engineering University 2011.
[6] Gong Hu, Zhuo Ying. Research on network situation awareness [J]. Software Journal 2010. 21
[7] Wei Yong, Lian Feng, Feng Dengguo. Network security situation assessment model based on information fusion [J]. computer research and development 2017.46 (3).