INTRUSION ALERT CORRELATION BASED ON D-S EVIDENCE THEORY

Mei Haibin, Gong Jian
School of Computer Science and Engineering, Southeast University
Computer Network Technology Key Laboratory of Jiangsu Province
Nanjing, China

Abstract—**Current intrusion detection systems (IDSs) often trigger a large amount of alerts, most of which are redundant alerts and false positives. Consequently, it is difficult for administrators to understand the alerts and take appropriate actions. Several alert correlation methods have been proposed. However, these methods don't consider the differences in reliability among alerts reported from multiple IDSs. This paper presents a novel alert correlation approach based on the Dempster-Shafer evidence theory, which regards the alerts as evidence of network attack and combines all the evidence according to the Dempster's combination rule, inferring whether the attack has taken place. The main advantage of the approach is that it can eliminate the ambiguity and confliction in alerts and reduce the number of alerts. With the DARPA 2000 test dataset, experimental results demonstrate that the approach can reduce more than 69% of reported alerts and decrease the false positive rate efficiently.**

Keywords-network security; intrsion detection system; alert correlation; D-S evidence theory

I. INTRODUCTION

Intrusion detection system (IDS) is considered as an effective second line of defense against attacks directed at computer systems [1]-[2]. In order to provide in-depth protection for network systems, administrators are increasingly deploying multiple IDSs. However, these IDSs often output a large amount of alerts, with lots of redundant alerts and false alerts in them, which overwhelms the analysis ability of security staff [3]-[7] and makes it a frustrating task to identify real alerts and find valuable attack information. Therefore, designing advanced alert correlation methods to reduce the redundancy of alerts and eliminate the false positives is increasingly important. Alert correlation has been an active research topic in the field of intrusion detection.

Recently, scholars have proposed some alert correlation methods. In [4]-[6], they propose an intrusion alert correlator based on the prerequisites of intrusions. An aggregation and correlation algorithm is presented in [8] for acquiring the alerts and relating them. Valdes [9] introduces a probabilistic approach for the coupled sensors to reduce the false alerts. In [10]-[12], a concept clustering method is used for finding the roots of false alerts and eliminating these roots to reduce the false alerts.

This research is supported by 973 Natural Key Basis Research and Development Plan of China (2003CB314804), and grant of Computer Network Technology Key Laboratory of Jiangsu Province (BM2003201).

However, none of them considers the difference in the reliability among alerts reported from diverse IDSs when alerts are correlated. They usually assume that all alerts have the same reliability. In fact, as IDSs use different detection techniques, they have different pros and cons. For example, one IDS can detect some network attacks with high accuracy, but may have problems in detecting other attacks. In addition, even the alerts that concern the same network attack, which are triggered by IDSs in different sites, may have different trustworthiness. Generally, alerts reported from sites far from the attack are less trustworthy than those reported from sites near to the attack. Therefore, the reliability of alerts from different IDSs is not equivalent, and this characteristic of alerts should be taken into account in alert correlation. If the difference in alerts reliability is fully considered, the uncertainty and ambiguity of alerts can be eliminated and the false positive rate will be decreased.

This paper presents a novel alert correlation approach, which deals with the different reliability of alerts by incorporating an alerts confidence fusion algorithm based on the D-S evidence theory. The approach regards the alerts reported from various IDSs as different evidence of the network attack, and assigns belief values to the evidence. Then, using the Dempster's combination rule, it fuses these belief values to infer whether the network attack reported by these alerts is true. With the D-S evidence theory, ambiguity and confliction in the alerts can be eliminated and the number of false alerts will be reduced. We conduct experiments on the DARPA2000 IDS test dataset and experimental results demonstrate that the proposed approach can reduce more than 69% of the alerts and efficiently decrease the false alert rate.

The rest of the paper is organized as follows. In the next section the fundamentals of the D-S theory are described. Section III elaborates on how to use the D-S theory to correlate alerts reported from various IDSs. Empirical results and comparison with related work are given in section IV. In the end section V concludes the whole paper and gives possible future work.

II. FOUDAMENTALS OF THE D-S THEORY

The D-S theory is a mathematical theory of evidence introduced in the 1960's by Arthur Dempster [13] and developed in the 1970's by Glenn Shafer. The D-S theory is appealing partly because it can handle uncertainty or ignorance,

that is, the lack of knowledge of the complete probabilistic model required for Bayesian inference. It has been applied to the fields of statistical inference, diagnostics, risk analysis and decision analysis [14].

The frame of discernment (FOD) in the D-S theory is a set of mutually exclusive and exhaustive possibilities denoted by Θ, which is similar to a state space in probability. A hypothesis H refers to a subset of Θ for which observers can represent evidence. In other words, H is an element in the power set of Θ.

Definition 1: *basic probability assignment (BPA)*. A basic probability assignment over Θ is a function m: $2^{\Theta} \rightarrow [0, 1]$, if and only if:

$$m(\phi) = 0 \ \ and \ \ \sum_{A \subseteq \Theta} m(A) = 1 . \tag{1}$$

The elements in 2^{Θ} associated to non-zero values of m are called focal elements and their union core.

Definition 2: *belief function (Bel)*. A function Bel: $2^{\Theta} \rightarrow [0, 1]$ over Θ is belief function such that:

$$Bel(A) = \sum_{B \subseteq A} m(B) . \tag{2}$$

The value Bel(A) quantifies the belief to proposition A.

Definition 3: *plausibility function (Pl)*. A function Pl: $2^{\Theta} \rightarrow [0, 1]$ over Θ is plausibility function such that:

$$Pl(A) = \sum_{B \cap A \neq \phi} m(B) . \tag{3}$$

Pl(A) is called the plausibility of A, which quantifies how we disbelieve A or how much A is not reliable.

Definition 4: *Dempster's rule of combination*. Dempster's rule of combination represents the conjunctive operation of the evidence. Let m_1 and m_2 be two *BPA*s over Θ, and their focal elements are $A_1, A_2, \ldots, A_m$ and $B_1, B_2, \ldots B_n$ respectively. If $\sum_{A_i \cap B_j = \phi} m(A_i) m(B_j) < 1$, then $m_{12} = m_1 \oplus m_2$ is defined as:

$$m_{12}(A) = \begin{cases} 0 & A = \phi \\ K^{-1} \sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j) & A \neq \phi \end{cases} . \tag{4}$$

where $A \subseteq \Theta$, and

$$K = 1 - \sum_{A_i \cap B_j = \phi} m_1(A_i) m_2(B_j) = \sum_{A_i \cap B_j \neq \phi} m_1(A_i) m_2(B_j) . \tag{5}$$

Correspondingly, let $m_1, m_2, \ldots, m_n$ be *BPA*s over Θ, the general Dempster's rule of combination $m_{1..n} = m_1 \oplus m_2 \oplus \ldots \oplus m_n$ is defined as [15]:

$$m_{1\ldots n}(A) = \begin{cases} 0 & A = \phi \\ K^{-1} \sum_{A_1 \cap \ldots \cap A_n = A} m_1(A_1) m_2(A_2) \ldots m_n(A_n) & A \neq \phi \end{cases} . \tag{6}$$

where $A \subseteq \Theta$, and

$$K = 1 - \sum_{A_1 \cap \ldots \cap A_n = \phi} m_1(A_1) m_2(A_2) \ldots m_n(A_n) = \sum_{A_1 \cap \ldots \cap A_n \neq \phi} m_1(A_1) m_2(A_2) \ldots m_n(A_n) . \tag{7}$$

III. ALERT CORRELATION APPROACH BASED ON THE D-S THEORY

A. The Framework of System

Based on the D-S theory, we propose an alert correlation method that fuses alerts generated by various IDSs. Fig. 1 shows the framework of IDS alert correlation system using our method. Here, raw alerts are first received and translated to the unified formal alerts in ARM (Alert Receive Module), and then redundant alerts are eliminated by REM (Redundance Elimination Moudle), which has the function of combining alerts according to some criteria.

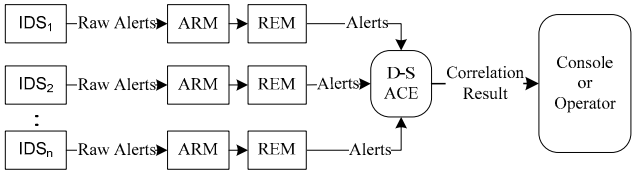

Figure 1. The framework of IDS alert correlation system based on the D-S theory.

The outputs of REMs are sent to the D-S ACE (Alert Correlation Engine), where alerts from multiple sources are combined by the Dempster's rule of combination, and the combination belief of input alerts will be calculated. According to the final combination belief value, the alerts with low belief value will be regarded as false positives and be filtered while alerts with high belief value will be deemed as true positives. In the end, the correlation results reported by D-S ACE will be presented to the console or operator.

B. The Unification Format of Alert

Because alerts generated by different IDSs have been encoded in dissimilar formats, it is necessary to translate each alert into a unification format understood by the alert correlation system. For implementing and testing our method easily, in this paper, we use a simple unification format instead of the IDMEF [16] alert format. Table I gives a brief description of the unification alert format.

TABLE I. THE GENERAL FIELDS OF ALERT

Field Name	Description
analyzerID	intrusion detection system ID
alertID	alert ID (mapping from Bugtraq ID,CERT ID and MITRE CVE)
alertType	class of alert
srcIP	source IP address of alert
dstIP	destination IP address of alert
srcPort	source port of alert
dstPort	destination port of alert
alertTime	alert time stamp

alertInfo	alert information and additional description

C. The Alert Correlation Engine

As mentioned before, we use alerts from several different IDSs as the evidence of hypothesis (for example, network attack occurrence).The function of ACE is to infer whether the attack has taken place with the D-S evidence theory. We first construct a BPA function for the evidence and calculate the belief assignment of evidence respectively for all hypotheses. Then, the Dempster's combination rule is used to combine these belief assignments and compute the total belief values. In the end, we use the total belief values and decision rule to estimate whether the attack has taken place. Fig. 2 shows the basic principle of the alert correlation engine, which consists of four steps as follows.

Step 1: Build the FOD for alert correlation

In order to decrease false positive rate, we first need to judge whether the network attack has really taken place by considering the alerts reported from various IDSs. If the network attack associated with these alerts has not happened, these alerts are considered as false alerts and need to be filtered. So the frame of discernment is defined as $\Theta = \{H, \neg H\}$, where H represents the hypothesis that the network attack associated with these alerts does take place, and $\neg H$ represents that the network attack doesn't take place. It clearly satisfies $H \cap \neg H = \varnothing$. Accordingly, the whole hypotheses are $2^{\Theta} = \{\varnothing, \{H\}, \{\neg H\}, \{H, \neg H\}\}$, where the hypothesis $\{H, \neg H\}$ means that it is unable to decide whether the attack takes place according to the received evidence at present.

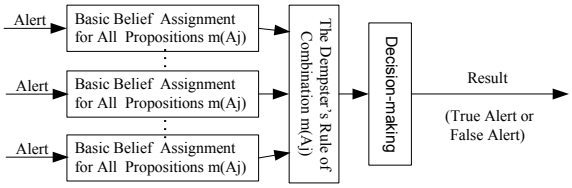

Figure 2. The principle of alert correlation engine based on the D-S theory.

Step 2: Choose the attack evidence

In this paper, IDSs are looked as the observers of network attacks. The reported alerts from each IDS are chosen as evidence of attack. For a certain network attack NA, there will be two possible values in reported alerts, namely, the network attack is reported in the alerts or is not reported. We use symbol R denotes being reported in alerts and $\neg R$ denotes not being reported.

Step 3: Design the BPA function m for evidence

In the D-S evidence theory, there isn't a general method for building the BPA function *m*. We design the BPA function *m* based on the characteristics of IDSs. In this paper, the BPA function *m* is defined in (8):

$$m: 2^{\Theta} \to [0,1], \quad m(\phi) = 0, \quad \sum_{I \in \{\{H\},\{\neg H\},\Theta\}} m(I) = 1. \tag{8}$$

where $m(\{H\})$ represents the belief value of current evidence which supports the hypothesis H, and $m(\{\neg H\})$ represents the belief value of current evidence which doesn't support the occurrence of network attack, while $m(\{H,\neg H\})$ represents the belief value of current evidence which supports the ignorance of network attack.

We set the value of BPA function according to two factors. One is IDS's different degrees of accuracy in detecting different types of network attacks; the other is the relationship between the position of ids and the attacked host. In general, if an IDS has a high accuracy for detecting the attack or it is near to the hosts which are under attack, the alerts generated by this IDS will have a high belief value of supporting the attack occurrence, and thus *m(H)* will be assigned a bigger value. In practical application environments, the first factor can be determined by statistical method or by experience of network security specialists, and the second factor usually relies on the deployed environment of IDSs and the hosts been monitored.

Step 4: Evidence combination and making decision

Suppose n is the number of IDSs involved monitoring a network attack NA, $\{m_1, m_2, \ldots, m_n\}$ is the set of basic belief assignment corresponding to evidence from every IDS. Combined belief assignment $m_{1..n}$ is computed by (9) and combination rule of two pieces of evidence is shown in (4) and (5).

$$\begin{aligned} m_{1\ldots n}(A) &= m_{1\ldots n-1}(A) \oplus m_n(A) \\ &= (m_{1\ldots n-2}(A) \oplus m_{n-1}(A)) \oplus m_n(A) \\ &\ldots\ldots \\ &= m_1(A) \oplus m_2(A) \oplus \ldots \oplus m_{n-1}(A) \oplus m_n(A) \end{aligned} \tag{9}$$

where $A \in \{\{H\},\{\neg H\},\{H, \neg H\}\}$.

When all evidence are combined, we can get the fusion belief results $m_{1\ldots n}(H)$, $m_{1\ldots n}(\neg H)$ and $m_{1\ldots n}(\Theta)$, respectively representing the final belief in supporting attack occurrence, not supporting and supporting the uncertainty of attack. We believe network attack doesn't occur, and consider the alerts associated with the attack as false positive when $m_{1\ldots n}(\neg H)$ is the biggest of the three values. Under other situations, we deem network attack occurs. Fig. 3 gives a detail description of our decision method.

If $\max(m_{1\ldots n}(H), m_{1\ldots n}(\neg H), m_{1\ldots n}(\Theta)) = m_{1\ldots n}(H)$
Result = true alert;
If $\max(m_{1\ldots n}(H), m_{1\ldots n}(\neg H), m_{1\ldots n}(\Theta)) = m_{1\ldots n}(\Theta)$
Result = true alert;
If $\max(m_{1\ldots n}(H), m_{1\ldots n}(\neg H), m_{1\ldots n}(\Theta)) = m_{1\ldots n}(\neg H)$
Result = false alert;

Figure 3. The rule of final decision.

IV. EXPERIMENTAL RESULTS

To test the performance of our proposed alert correlation approach on its ability to reduce alert number and false positive rate, we conduct a series of experiments using the DARPA

2000 intrusion detection evaluation datasets [17], which consist of two datasets: LLDOS 1.0 and LLDOS 2.0. Each dataset includes network traffic data collected from both the DMZ and the inside part of the evaluation network. They include two multi-stage distributed DoS attack scenarios. In every scenario, the attack consists of five phases: probing the network, breaking into those hosts with the Solaris *Sadmind* vulnerability, installing the DDoS software on the compromised hosts, and launching a DDoS attack at a remote server.

In the experiments, the LLDoS 1.0 dataset is used as our testing data. One alert set is generated by the RealSecure Network Sensor 6.0 [18] on the LLDDoS 1.0 DMZ data, whose configuration is the rule of maximum coverage. Another alert set is generated by Snort 2.0 –an open-source network-based IDS– on the LLDDoS 1.0 inside data, which uses the up-to-date rules [19]. Fig. 4 and Fig. 5 illustrate the alert type and number in the two alert sets. Snort reports 28 kinds of alerts (see Fig. 4), while Realsecure Network Sensor reports 20 categories (see Fig. 5).

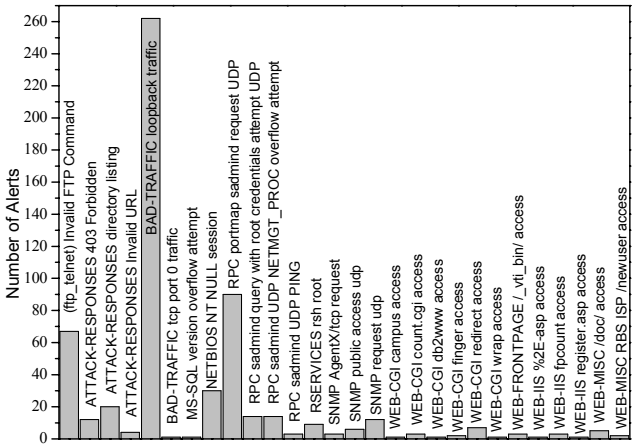

Figure 4. The reported results of Snort.

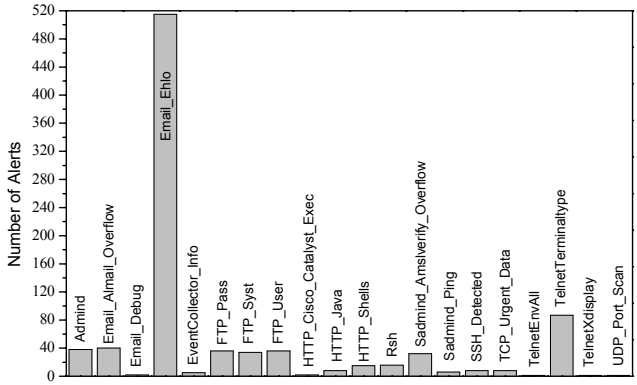

Figure 5. The reported results of Realsecure Network Sensor.

Table II summaries the number of alerts and the false positive rate reported by the above two IDS systems using LLDDoS 1.0 dataset. In Table II, the false positive rate is defined as:

$$FPR = \frac{NOFA}{NOA} \times \% = \frac{NOA - NOTA}{NOA} \times \% . \qquad (10)$$

where *FPR* is the false positive rate, *NOFA* is the number of false positives, *NOA* is the number of alerts, and the *NOTA* is the number of true alerts.

TABLE II. THE NUMBER OF ALERTS AND THE FALSE POSITIVE RATE

Dataset	IDS	Number of alerts	Number of true alerts	FPR
LLDOS1.0 DMZ	RealSecure Network Sensor 6.0	891	57	93.60%
LLDOS1.0 Inside	Snort 2.0	840	44	94.76%

In the raw alerts reported from IDS, there will contain multiple redundant alerts generated by an attack. Therefore, in the alert correlation system, we implement an ARM to eliminate the redundancy alerts of every IDS (as shown in Fig. 1) before fusing alerts from various IDSs based on D-S theory. The method of redundancy elimination is discussed at length in [20]. Table III shows the output results of ARMs. Then, we adopt the new alert correlation method based on D-S theory to correlate alerts from the two ARMs. The experimental result is listed in Table IV.

TABLE III. THE NUMBER OF ALERTS AND THE FALSE POSITIVE RATE AFTER REDUNDANCY ELIMINATION

Dataset	IDS	Number of alerts	Number of true alerts	FPR
LLDOS1.0 DMZ	RealSecure Network Sensor 6.0	113	24	78.76%
LLDOS1.0 Inside	Snort 2.0	95	21	77.89%

TABLE IV. THE NUMBER OF ALERTS AND THE FALSE POSITIVE RATE AFTER CORRELATION WITH PROPOSED METHOD

Dataset	Number of alerts	Number of true alerts	FPR
LLDOS1.0	64	34	46.88%

As shown in Table III and Table IV, our alert correlation algorithm based on D-S theory can further reduce the number of alerts reported from the two ARMs and decrease the false positive rate. After correlated, the number of reported alerts are reduced from 208(113+ 95) to 64. The decreased number of alerts is about 69 percent of the output alerts of ARMs. The false positive rate is also decreased from 78.33% ((78.76% + 77.89%)/2) to 46.88%. Compared with another representative alert correlation method which uses the prerequisites and consequences of attack, the time complexity of our method is only O(n), for most computation spends on the combination of evidence. Moreover, the proposed approach doesn't require maintaining large amounts of intermediate correlation states so that it is much easier to be applied to online implementation.

V. CONCLUSION

In this paper, we propose a novel alert correlation method which implements the function of alerts confidence fusion. The approach is based on the observation that alerts from heterogeneous IDSs have different trustworthiness and should

not be treated equally. Based on the D-S evidence theory, we represent the alerts from various IDSs as attack evidence and assign them different beliefs. By fusing the evidence, we can infer whether the network attack has really taken place and achieve the goal of improving the quality of IDS alerts. We perform experiments on the DARPA 2000 intrusion detection evaluation datasets. Experimental results show that, with alerts confidence fusion, our alert correlation method can potentially eliminate the ambiguity and confliction in the alerts reported from different IDSs, and further decrease the number of alerts and reduce the false positive rate. In the future, our work will focus on the setting up of more delicate BPA functions and examine the D-S evidence theory in greater depth.

ACKNOWLEDGMENT

The authors would like to thank anonymous reviewers for their perspicacious comments and valuable advice.

REFERENCES

[1] R. Bace, Intrusion detection. Indianapolis: Macmillan Technology Publishing, 2000.

[2] D. E. Denning, "An intrusion detection model," IEEE Transactions on Software Engineering, vol.13, pp. 222-232, 1987.

[3] J. Li and Z. Li, "Correlation analysis for distributed intrusion alert," Journal of Computer Research and Development, vol. 41, pp.1919-1923, 2004.

[4] P. Ning, Y. Cui, D. S. Reeves, and D. Xu, "Tools and techniques for analyzing intrusion alerts," ACM Trans. on Info. and System Security, vol. 7, no. 2, pp. 273-318, 2004.

[5] P. Ning, D. Xu, C. G. Healey, and R. St. Amant, "Building attack scenarios through integration of complementary alert correlation methods," In the Proc. of the 11th Annual Network and Distributed System Security Symposium. San Diego, pp. 97-111, 2004..

[6] P. Ning and D. Xu, "Learning attack strategies from intrusion alerts," In Proc. of the 10th ACM Conference on Computer and Communications Security. Washington D.C., pp. 200-209, 2003.

[7] H. Debar and A. Wespi, "Aggregation and correlation of intrusion detection alerts," In Proc. of 4th International Symposium on Recent Advances in Intrusion Detection (RAID). Davis, pp. 85-103, 2001..

[8] P. Roberto, G. Giorgio, and R. Fabio, "Alarm clustering for intrusion detection systems in computer networks," Engineering Applications of Artificial Intelligence. vol.19, pp. 429–438, 2006.

[9] A. Valdes and K. Skinner, "Probabilistic alert correlation," In Proc. of the 4th International Symposium on Recent Advances in Intrusion Detection (RAID). Davis, pp. 54-68, 2001.

[10] K. Julisch, "Clustering intrusion detection alarms to support root cause analysis," ACM Trans. on Information and System Security. vol. 4, no. 6, pp. 443-471, 2003.

[11] K. Julisch and M. Dacier, "Mining intrusion detection alarms for actionable knowledge," In Proc. of the 8th International Conference on Knowledge Discovery and Data Mining. New York, pp. 366-375, 2002.

[12] K. Julisch, "Mining alarm clusters to improve alarm handling efficiency," In 17th Annual Computer Security Applications Conference. New Orleans, pp. 12-21, 2001.

[13] A. Dempster, "Upper and lower probabilities induced by multivalued mapping," Annals of Mathematical Statistics. vol. 38, no. 2, pp. 325−339, 1967.

[14] J. Zhuge, D. Wang, Y. Chen, Z. Ye, and W. Zou, "A network anomaly detector based on the D-S evidence theory," Journal of Software. vol. 17, no. 3, pp.463-471, 2006.

[15] Y. Kang, Theory and application of data fusion. Xian: Press of Electronic Technology University. 1997.

[16] D.Curry and Hervé Debar, Intrusion detection message exchange format data model and extensible markup language (xml) document type definition1 IETF. http://www.ietf.org/lid-abstracts.html, 2003.

[17] MIT Lincoln Lab, 2000 DARPA intrusion detection scenario specific dataset. http://www.ll.mit.edu/IST/ideval/data/2000/2000_data_index.html, 2003. 07.

[18] ISS, Inc.: RealSecure intrusion detection system. http://www.iss.net.

[19] M. Roesch, "Snort – lightweight intrusion detection for network," In Proc. of the 13th System Administration Conference, LISA 1999, www.snort.org/docs/lisapaper.txt.

[20] J. Gong, H. Mei, Y. Ding, and D. Wei, "A multi-feature correlation redundance elimination of intrusion event," Journal of Southeast University. vol. 35, no. 3, pp. 366-371, 2005.