Research on situation awareness of network security assessment based on dempster-shafer

Weifa Zheng*

21 Chisha Road, Haizhu District, Guangzhou, Guangdong, China

Keywords: Network security, Situation assessment, Dempster-shafer.

Abstract. There are a lot of uncertainties in the network security situation assessment that depends on is multi-source and heterogeneous. Therefore, the objective uncertainties must be described and dealt with in the process of network security situation assessment. This paper proposes a multi-attribute decision-making trust evaluation model based on D-S evidence theory in multi-source and heterogeneous environment. By collecting, processing and evaluating attack event information from many data sources of security device, it can effectively evaluate the possibility of network intrusion, and provide a new exploration for network security situation assessment.

1 Introduction

With the rapid development of computer and communication technology, network attacks are gradually developing towards the trend of distribution, scale, complexity, etc. A single firewall, such as intrusion detection, anti-virus, access control or other network security technologies, can't meet the requirements of network security. Therefore, there is an urgent need for new technologies to sense the abnormal events in the early warning network, to get hold of the network security situation in real time, and to improve the network security protection ability. [1] The network security situation awareness can fully perceive the network security threat situation, understand the health status of the network and application. It can help security personnel take targeted response measures.[2] Under the environment of the big data, network security situation awareness has the characteristics of multiple sensing nodes, diverse data types, dynamic information and so on. It contains a lot of uncertain information, which is incomplete, imprecise and contradictory to a certain extent. It needs to solve the problem of uncertain information in the situation assessment process.

D-S evidence theory is originated from the work of mathematician A. P. Dempster of Harvard University in 1960s, which is in using upper and lower limit probability to solve the problem of multi-valued mapping. G. Shafer, a student of Dempster, made further development of D-S evidence theory by introducing introduced the concept of trust function, and formed a set of mathematical methods based on "evidence" and "combination" to deal with uncertain reasoning problems. In the complex network, the data that network security situation assessment depends on is multi-source, heterogeneous, and uncertain. Therefore,

* Corresponding author: wf@gdufe.edu.cn

the objective uncertainties must be described and dealt with in the process of network security situation assessment. Based on the combination of intrusion detection features and evidence theory, aiming at the characterization of various intrusion behaviors in the network, this paper uses D-S evidence theory to evaluate the credibility of the intrusion threat, which provides a new method for the security situation assessment of the network.

2 Network security situation assessment

The concept of situation awareness (SA) was put forward by Endsley in 1988. SA is the acquisition, understanding and short-term prediction of environmental factors in a certain time and space. The purpose of network situation awareness is to display and predict the latest development trend of the network by acquiring, understanding the equipment running status, network behavior and user behavior that can cause the change of network situation in large-scale network environment. Network security situation awareness refers to the real-time collection of all data flows in the network, the analysis and processing of the flows by using data fusion and data mining technology, and the visual display of big data technology to show the operation security of the network in real time, so as to provide guarantee for network security [3-5]. In order to detect potential and malicious attacks in real time and accurately, network security situation awareness should include data preprocessing, feature extraction, situation assessment, situation prediction and situation display. [6-7]

The data of network security situation awareness comes from a large number of heterogeneous hosts and networks, including firewalls, web application firewalls, intrusion detection systems(IDS), intrusion prevention systems(IPS), database audit equipment, etc. The quantity of security log data generated by these security devices are growing rapidly. Different kinds of log data contain a large amount of uncertain information, which is incomplete, inaccurate and contradictory to some extent, so it is necessary to solve the problem of uncertain information in the situation assessment process by integrating the complementary, redundant and even contradictory data.

3 D-S evidence theory

When G.Shafer published the book *A Mathematical found of Evidence* in 1976, it marks the Evidence Theory became a complete theory of dealing with uncertainty. In D-S evidence theory, the recognition framework Θ is used to represent the complete set of the object studied, and the elements in Θ are mutually exclusive and discrete values. The set of all subsets based on the recognition framework Θ is called the power set of Θ , which is called 2^Θ . [8][9][10]

Definition 1: Basic Probability Assignment Function m (BPA). In the identification framework Θ , if $m: 2^\Theta \rightarrow [0, 1]$ exists, it is satisfied

$$\begin{cases} m(\emptyset) = 0 \\ \sum_{A \subseteq \Theta} m(A) = 1 \end{cases} \quad (1)$$

Then m is the basic probability distribution function of A , called mass function, so that A of $m(A) > 0$ is called Focal elements.

Definition 2: Belief function. The Belief function based on BPA m on the identification framework Θ is defined as:

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

Definition 3: Plausibility function, the Plausibility function based on BPA on the identification framework Θ is defined as follows:

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B) \quad (3)$$

Definition 4: Reliability interval. The Belief function $Bel(A)$ and the Plausibility function $Pl(A)$ constitute the reliability interval $[Bel(A), Pl(A)]$, which is used to express the degree of confirmation of the proposition. $[0,0]$ represents the negation of proposition A, $[1,1]$ represents the affirmation of proposition A. The description of the evidence reliability interval is shown in figure 1 below:

Fig. 1. D-S reliability interval description.

Definition 5: Dempster's combinational rule, also known as evidence combination formula. For two mass functions m_1 and m_2 on $\forall A \subseteq \Theta$ and Θ , Dempster's combination rule is:

$$m_1 \oplus m_2(A) = \frac{1}{K} \sum_{B \cap C = A} m_1(B) \cdot m_2(C) \quad (4)$$

where, K is the normalized constant

$$K = \sum_{B \cap C \neq \emptyset} m_1(B) \cdot m_2(C) = 1 - \sum_{B \cap C = \emptyset} m_1(B) \cdot m_2(C) \quad (5)$$

For $\forall A \subseteq \Theta$, identify the finite mass functions $m_1, m_2, \dots, m_n$'s Dempster combination rule is:

$$(m_1 \oplus m_2 \oplus \dots \oplus m_n)(A) = \frac{1}{K} \sum_{A_1 \cap A_2 \cap \dots \cap A_n = A} m_1(A_1) \cdot m_2(A_2) \cdots m_n(A_n) \quad (6)$$

where

$$\begin{aligned} K &= \sum_{A_1 \cap \dots \cap A_n \neq \emptyset} m_1(A_1) \cdot m_2(A_2) \cdots m_n(A_n) \\ &= 1 - \sum_{A_1 \cap \dots \cap A_n = \emptyset} m_1(A_1) \cdot m_2(A_2) \cdots m_n(A_n) \end{aligned} \quad (7)$$

K known as the normalization factor, reflect the degree of evidence conflict between, when K is close to zero, the smaller the degree of the conflict between body of evidence, the more accurate-fusion results, when K is close to 1, the greater the degree of the conflict between body of evidence, the fusion result is not accurate, when K = 1, the contradiction between the evidence body, can't carry on the effective fusion.

Definition 6: class of probability functions.

$$f(A) = Bel(A) + \frac{|A|}{|\Theta|} \times (Pl(A) - Bel(A)) \quad (8)$$

where $|A|$ and $|\Theta|$ are the number of elements in A and Θ respectively. The certainty of A proposition can be expressed by the quasi-probability function.

4 Network security situation evaluation model based on D-S evidence

In this paper, the D-S evidence theory is used to fuse the security data sources. By dealing

with the characteristic attributes of network security events, the basic probability distribution is established, the modified BPA function can be obtained by weight calculation, and the credibility of network attacks is evaluated according to the synthesis rules.

In this paper, T_e is the credibility of the network attack event. The credibility of the attack event is measured by multi-dimensional characteristic attributes, including four characteristic attributes. In the trust model of this paper, Θ is defined as $\{\text{Attack}, \text{NO Attack}\}$. The relationship between intruder and event is divided into trust $\{\text{Attack}\}$, distrust $\{\text{NO Attack}\}$, uncertainty $\{\text{Attack}, \text{NO Attack}\}$. Among them, S1 represents the degree of threat calculated by the logs of firewall, S2 represents the degree of threat calculated by the logs of web application, S3 represents the degree of threat calculated by the logs of IDS.

Event basic probability distribution. The basic probability distribution is shown in the following table:

Table 1. Basic probability distribution table.

Result	S1	S2	S3
$\{\text{Attack}\}$	m_{11}	m_{21}	m_{31}
$\{\text{NO Attack}\}$	m_{12}	m_{22}	m_{32}
$\{\text{Attack}, \text{NO Attack}\}$	m_{13}	m_{23}	m_{33}

The basic probability distribution functions of S1,S2 and S3 in the identification framework Θ are m_1, m_2, m_3 respectively.

Based on the basic probability distribution corresponding to each evidence, the importance weight of evidence is calculated by using equations (1) ~ (7)

According to formula (4), attribute evidence is synthesized to obtain trust triples $\{m(\{\text{Attack}\}), m(\{\text{NO Attack}\}), m(\{\text{Attack}, \text{NO Attack}\})\}$ representing trust of attack event.

Formula (8) class probability function is used to calculate the trust value. As a measure of the imprecision of trust, the reliability space of uncertain events is divided according to the probability of occurrence of trusted and untrusted events in the data source, and the two events are assigned to be trusted and untrusted.

$$T_e = f(\text{Attack}) = \text{Bel}(\text{Attack}) + \frac{|\text{Attack}|}{|\{\text{Attack}, \text{NO Attack}\}|} \times (Pl(\text{Attack}) - \text{Bel}(\text{Attack})) \quad (9)$$

$|\text{Attack}| = 1, |\{\text{Attack}, \text{NO Attack}\}| = 2$. Formula (9) comprehensively considers the trusted part, the untrusted part and the uncertain part of trust relationship. This method can accurately evaluate the direct trust relationship between the security logs.

5 Experiments

In this paper, the network security situation assessment uses the log generated by three security devices to calculate the credibility of the attack event, in which S1 represents the degree of threat calculated by the firewall log record, S2 represents the degree of threat calculated by the web application firewall log record, S3 represents the degree of threat calculated by the intrusion detection system (IDS) log record. Taking S1 as an example, the value range is 0-1, which is calculated according to the connection duration, protocol type, and network service type of the target host in the firewall log. Randomly generate data of 100 attack events, select 10 of them, and the BPA is

Table 2. Basic probability assignment table.

NO	S1			S2			S3		
	Attack	NO Attack	Uncertainty	Attack	NO Attack	Uncertainty	Attack	NO Attack	Uncertainty
1	0.365388	0.507689	0.126922	0.604338	0.3165297	0.079132435	0.89697	0.082424	0.020606
11	0.863579	0.109137	0.027284	0.773456	0.1812354	0.045308858	0.946923	0.042461	0.010615
21	0.166659	0.666673	0.166668	0.614528	0.3083778	0.077094441	0.761412	0.19087	0.047718
31	0.696684	0.242653	0.060663	0.822007	0.1423946	0.035598654	0.504512	0.39639	0.099098
41	0.405911	0.475271	0.118818	0.619408	0.3044736	0.076118396	0.45007	0.439944	0.109986
51	0.712805	0.229756	0.057439	0.646673	0.2826613	0.070665316	0.046494	0.762805	0.190701
61	0.271219	0.583025	0.145756	0.313934	0.5488527	0.137213182	0.113324	0.70934	0.177335
71	0.771219	0.183025	0.045756	0.813934	0.1488527	0.037213182	0.813324	0.14934	0.037335
81	0.898288	0.08137	0.020342	0.70784	0.2337278	0.058431946	0.436024	0.451181	0.112795
91	0.198922	0.640862	0.160216	0.548133	0.3614936	0.090373403	0.581008	0.335193	0.083798

According to formula (1) ~ (9), We calculate the credibility of the above 10 attacks based on three different data sources which are collected from 3 security devices.

Table 3. Credibility of attack table.

NO	$m'_i(Attack)$	$m'_i(NO\ Attack)$	$m'_i(Attack, NO\ Attack)$	$Bel(Attack)$	$f(Attack)$	Result
1	0.923371	0.076619	0.000009	0.923371	0.923376	Attack
11	0.997737	0.002263	0.000000	0.997737	0.997737	Attack
21	0.706358	0.293562	0.000080	0.706358	0.706398	Attack
31	0.938127	0.061857	0.000016	0.938127	0.938135	Attack
41	0.624622	0.375144	0.000234	0.624622	0.624739	Attack
51	0.581553	0.418027	0.000420	0.581553	0.581763	Attack
61	0.103313	0.895076	0.001611	0.103313	0.104118	NO Attack
71	0.987192	0.012807	0.000001	0.987192	0.987193	Attack
81	0.959720	0.040260	0.000020	0.959720	0.959730	Attack
91	0.496755	0.502962	0.000283	0.496755	0.496896	NO Attack

6 Conclusion

This paper proposes a multi-attribute decision-making trust evaluation model based on D-S evidence theory in multi-source heterogeneous environment. It evaluates the credibility of network attack events by objective evidence, thus providing a new exploration for network situation evaluation technology. The paper proves that we can effectively evaluate the possibility of network intrusion by collecting, processing and evaluating the attack event information of multiple security device data sources. This research has certain reference significance for the development of network security situation.

References

1. Dong-Lan LIU, Xin LIU, Hao ZHANG, Hao YU, Lei MA, Xiao-Hong ZHAO. Research and Application of Network Security Situation Awareness and Active Defense Based On Big Data Technology[J].Computer Measurement &Control.2019,27(10); pp.229
2. Lenders V, Tanner A, Blarer A.Gaining an edge in cyberspace with advanced situational awareness[J].Security & Privacy IEEE,2015,13(2):65-74.
3. Xiao-Wu LIU, Hui-Qiang WANG, Hong-Wu LÜ, Ji-Guo YU, Shu-Wen ZHANG. Fusion-Based Cognitive Awareness-Control Model for Network Security Situation[J].Journal of Software,2016,27(8):2099-2114
4. Hui-Qiang WANG, Ji-Bao LAI, Liang ZHU, Ying LIANG. Survey of Network Situation Awareness System [J].Computer Science, 2006, 33(10):5-10
5. Zheng-Hu GONG, Ying ZHUO, Research on Cyberspace Situational Awareness [J]. Journal of Software, 2010, 21(7):1605-1619.
6. Xu G , Cao Y , Ren Y , et al. Network Security Situation Awareness Based on Semantic Ontology and User-defined Rules for Internet of Things[J]. IEEE Access, 2017:1-1.
7. Xiao-Lu HAN, Review of the Theory, Technology and Some Difficult Problems in Network Security Situational Awareness[J], Information Security and Communications Privacy, 2019,7(7):67
8. Dempster, A. P. Upper and lower probabilities induced by a multivalued mapping. Annals of Mathematical Statistics, 1967, 38(2): 325-339.
9. Dempster, A. P. Generalization of Bayesian Inference. Journal of the Royal Statistical Society. Series B 30, 1968:205-247.
10. Shafer, G. A Mathematical Theory of Evidence. Princeton University Press, 1976.