# A Study of Various Intrusion Detection Model Based on Data Fusion, Neural Network and D-S Theory

**Ramnaresh Sharma[1], Manish Shrivastava[2]**
PG Scholar LNCT, Bhopal (M.P)[1]
Head of Information Technology Bhopal (M.P.) Department, LNCT, Bhopal (M.P.)[2]

## Abstract

***Network security and awareness of network attack are hot pots in current research area. Now in days various model and method are available for intrusion detection and awareness of cyber-attack. Such as Application of the integrated Network Security Situation Awareness system (Net-SSA) shows that the proposed framework supports for the accurate modeling and effective generation of network security situation. In this paper we have discuss various approach for intrusion detection technique such as data fusion, neural network and D-S Theory and fuzzy logic.***

## Keywords

***Intrusion Detection, Data Fusion, Neural Network***

## 1. Introduction

Traditional network security devices such as Intrusion Detection Systems (IDS), firewalls, and security scanners operate independently of one another, with virtually no knowledge of the network assets they are defending. This lack of information results in numerous ambiguities when interpreting alerts and making decisions on adequate responses. Network systems are suffering from various security threats including network worms, large scale network attacks, etc, and network security situation awareness is an effective way for solve these problems. The general process is to perceive the network security events happened in a certain time period and cyberspace environment, synthetically manipulate the security data, analyze the attack behaviors systems suffered, provide the global view of network security, and assess the whole security situation and predict the future security trends of the network. There exist several difficulties when implementing network security situation awareness. [1] The amount of alert events generated from various security sensors is tremendous and the false positive rate is too high. [2] The trivial alerts generated from large scale network attacks (e.g. DDoS) are very complex and the relationships among them are difficult to determine. [3] The data type of alert events generated from security sensors are very abundant, while there is a lack of knowledge needed by alert processing, and automatically acquiring these knowledge is rather difficult after being put forward in the last century and decades of rapid development, data fusion technology has been widely used in the military, geological and chemical industry. And that multi-sensor data fusion has become an important method to analyze the large scale of heterogeneous data in the complex systems. In the multisensor data fusion system, the multiple sensors can gain more targets' information, and using these information and data appropriately can improve the system's measurement accuracy, enhance the fault-tolerance, improve its stability and reliability, and ultimately improve the system's overall performance. Therefore, in this paper, multi-sensor data fusion technology is introduced to the intrusion detection research, so that the results of different detecting methods and heterogeneous data in the system can be "fused together. Furthermore, the intrusion scene and the system's security situation can be extracted at a higher level.

The rest of this paper is organized as follows. In section 2 network situation elements. Section 3 gives a data fusion neural network and D-S Theory. Section 4 describes the intrusion model. Section 5 concludes this paper

## 2. Network Situation Element

Situational awareness was defined by Endsley as "the perception of the elements in the environment with a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future". Currently, there is no uniform and general definition of network situation awareness. Network situation indicates that the whole network current status and its changing trend according to some factors of running status of network facilities, network and user behavior, etc. Namely, situation is a status, a trend and a whole notion

### 2.1 Constitution of Network Security Situation Element

Network situation elements consist of Internet/Intranet (environment), entities in the network including software and hardware, network security events including alerts, logs and files, correlation team and network intrusion behavior. Their relationship is illustrated in Figure 1

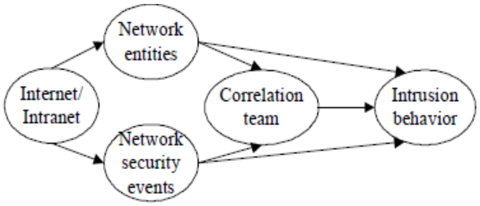

**Figure.1: Network situation element**

There are many entities in the environment of Internet/Intranet. When their situation changed, some events occurred. There are correlations in time and space between events occurring on each entity. The combining of several entities and events according to certain relationship will become a unit as correlation team. Correlation team or entity will cause some intrusion behavior. NSSA describes this process. When intrusion behavior occurs, not only security events are happening from IDS alerts, but also system state is changing during running time according to logs.

### 2.2 Concrete factors in the process of Security situation analysis

NSSA fuses data from tools of IDS, VDS (Virus Detection System), Firewall, Netflow etc. to find what happening in the network, whether there are intrusion events occurring, and predict the next goal. Security situation analysis involves in amount of factors in the network, as shown in Figure 2. However, as Lambert thought, the main aim of situation estimation is to describe the interested problem in the environment. Therefore, in the actual processing, several factors are needed. For instance, the sensors can be classified in two different types [3]: a preprocessor plugin for Snort and a SNMP data collector and analyzer; Y. Zhai used Snort and Norton antivirus to reason about intrusion evidence [4]

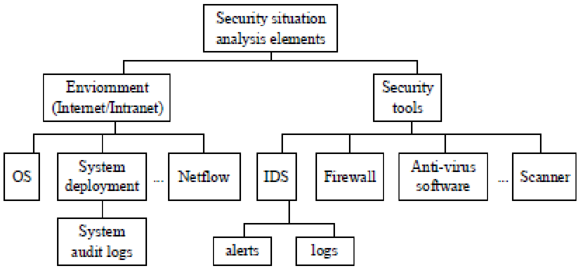

**Figure.2: Factors of affecting network security situation**

## 3. Overview of Data fusion, Neural Network and D-S Theory

Data fusion techniques combine data from different sources together. The main objective of employing fusion is to produce a fused result that provides the most detailed and reliable Information possible. Fusing multiple information sources together also produces a more efficient representation of the data Dempster-Shafer evidence theory.

The D-S framework is based on the view that proposition can he regarded as subsets of a given set of hypotheses. For example, in the intrusion detection system, we can regard the set of hypotheses as the set of categories of intrusion. Each anomalous event, then, is a subset of Θ. Thus, the propositions of interest are in a one-to-one correspondence with the subsets of Θ, and the set of all propositions corresponds to the set of all subset of Θ, which is denoted 2Θ. Θ is named a frame of discernment, and the proposition are said to be discerned by the frame A.

**Definition 1**: Basic Probability Assignment Beliefs can be assigned to propositions to express their uncertainty. The beliefs are usually computed based on a density function $m : 2\Theta \rightarrow [0,1]$ called a basic probability assignment(bpa) or mass function: $\Sigma\{m(A) \mid A \subseteq \Theta\} = 1 \; m(\varphi) = 0$ (1)
m(A) represents the belief exactly committed to A.

**Definition 2**: Belief Function Given a body of evidence with bpa m, we can compute the total belief provided by the body of evidence for a proposition. This can be done by a belief function $Bel: 2\Theta \rightarrow [0, 1]$ conjunctive operation of the evidence. Given several belief functions on the same frame of discernment based on the different evidences, if they are not entirely conflict, we can calculate a belief function using Dempster's rule of combination. It is called the orthogonal sum of the several belief functions. The orthogonal sum Bel1⊕Bel2 of two belief functions is a function whose focal elements are all the possible intersections between the combining focal elements and whose bpa is given by

$$m(A) = \frac{\sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j)}{1 - \sum_{A_i \cap B_j = \phi} m_1(A_i) m_2(B_j)} = \frac{1}{N} \sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j),\ A \neq \phi$$

$$m(A) = 0,\ A = \phi$$

$$N = 1 - \sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j) > 0 \tag{3}$$

In recent years, significant attention has been focused on multi-sensor data fusion for both military and non-military applications [4][5][7]. Data fusion techniques combine data from multiple sensors to achieve more specific inferences than could be achieved by using a single, independent sensor. Data fusion is a critical part of the NSSA. It is the process of collecting the information from multiple and possibly heterogeneous cyberspace sources and combining it in order to get a more descriptive, intuitive and meaningful result [3]. In a security system, there are multiple network security sensors, but it is often difficult to obtain a panoramic, allencompassing view of an overall situation of the security status of a complex system. This is especially true of systems that, for example, may be geographically distributed over a wide area. So it is extremely important to fuse the outputs of these sensors in an effective and intelligent manner in order to improve the robustness of the system and provide the analysts with an overall network security situation. Furthermore, multiple data sources can provide more robust performance due to the inherent redundancy. Therefore, data fusion techniques of combining data from several data sources can yield higher accuracy and robustness than that achieved by a single data source. As to fusion techniques, there are: the Bayesian theory, Neural Networks (NN), Support Vector Machines (SVMs) and so on. As a elementary research, we adopt the MLF-NN method as the fusion technique and it has the ability to deal with non-linear and multi-classification issues An MLF-NN is an information processing system that is inspired by the way biological nervous systems, such as the brain, process information. It is composed of a large number of highly interconnected processing elements (neurons) working with each other to solve specific problems. Each processing element (a neuron) is basically a summing element followed by an activation function. The output of each neuron is fed as the input to all of the neurons in the next layer. The learning process is essentially an optimization process in which the parameters of the best set of connection coefficients (weights) for solving a problem are found. We show the structure of a MLF-NN, which has two hidden layers in Figure 3.

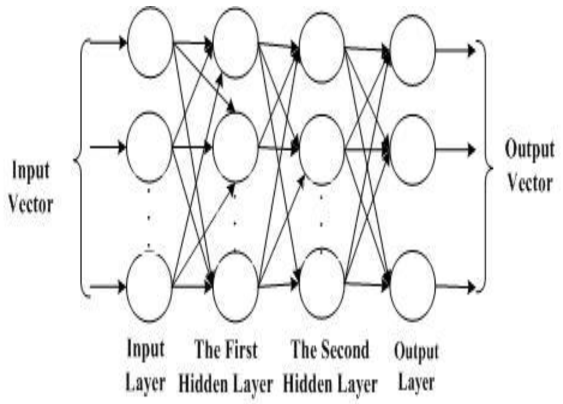

**Figure. 3: MLF-NN structure**

## 4. Security Model

**Network Security Situation Generation Model**

The NSSA Estimation (NSSAE) produces the security situation and the security risk of the current system according to the results of the fusion engine and the historical data stored in the Network Security Situation Database (NSS-DB). In an alert aggregation sub-module, we adopt an effective alert aggregation algorithm. First, we set a time window and summarize the number of alerts if their SIP (source IP address) and ATK _ TYPE (attack type) are equivalent to each other. After that, the alert aggression sub-module produces a vector, V(t) = (COUNT, ATK _TYPE, SIP, , ATK _ TYPE ) TIME PSeverity . COUNT represents the count of attacks and TIME indicates the identifier of the time window. What we call threaten gene denotes the degree of the severity of this kind of anomalous behavior. In most situations, the gene Pn (n {DoS obe U R R L}) Severity , Pr , 2 , 2 is determined according to the experience of the administrator, which is very coarse and inaccurate. In our framework, we adopt the weight-gene distribution approach described in [9] and introduce it into the field of network security as follows.

$$P^{n}_{Severity} = \begin{cases} \frac{1}{2} + \frac{\sqrt{-2\ln\frac{2i}{n}}}{6}, 1 \le i \le \frac{1}{2} \\ \frac{1}{2}, i = \frac{n}{2} \\ \frac{1}{2} - \frac{\sqrt{-2\ln\left(2-\frac{2i}{n}\right)}}{6}, \frac{n}{2} \le i \le n \end{cases}$$

*i* denotes the severity-rank number and *n* represents the classification number. These *V*(*t*) are recorded in the NSS-DB. We can calculate the current situation, *S* , in the current time window using (10) and (11),according to the *V*(*t*) s in this time window: =Σ { }

*n S Sn* , *n Dos*, Pr *obe*,*U*2*R*, *R*2*L*

(10)

where

*S COUNTn* {*DoS obeU R R L*}

*n ATK TYPE*

*P*

*n*

*n*

Here, we employ n PSeverity 10 to emphasize the importance of the degree of the severity of a given attack. For example, there are two types of attacks, DoS and U2R, and the values of theirCOUNT are 10 and 20, respectively. Then we calculate the value of their situation using (12) and of threaten genes shown All the values of S and Sn are stored in NSS-DB in order to search for historical security situation in the future.

### B. Hierarchical Model of the Data Fusion Based Intrusion Detection System

A 3-level structure which is generally accepted by scholars [1] is adopted in this paper. The whole system contains three layers: the basic detection layer, information layer and knowledge layer (as shown in the Fig. ).

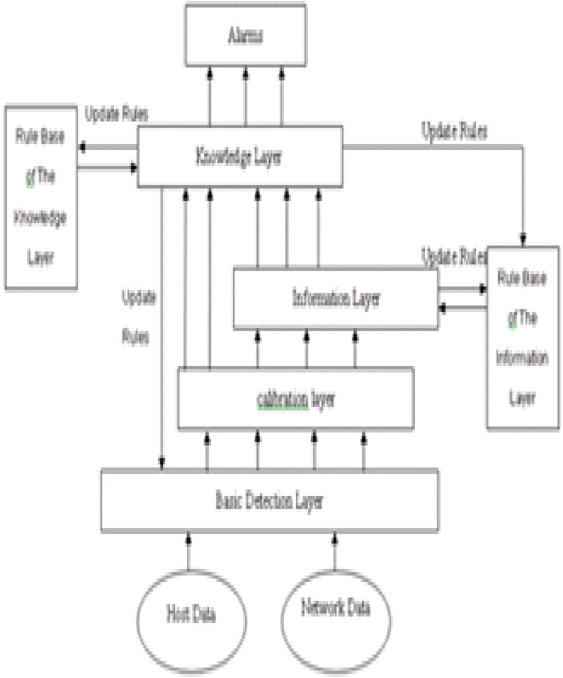

**Figure .4: Hierarchical model of the data-fusion based Intrusion detection system**

**(a)** The basic detection layer Various IDS agents are arranged in the basic detection layer, different agents adopt different detecting methods to give detection to the collected system information. Each detecting agent can be an independence intrusion detection system (For example: Snort), one kind of detection method (For example: SVM), or other computer security system (For example: Firewall). And, all the detecting results are transmitted to the data fuse module of upper layer to carry out the fusing. Though, each agent can only make a partial judgment to the system's security situation, combining the detecting information of all basic agents can provides sufficient and all-round safety information for the upper layer's fusing module, which is because the focus points of each kind of basic detecting agents are different ( For example: some agents adopt misuses detection but others adopt anomaly detection, or, some agents adopt host intrusion detection but others adopt network-based intrusion detection).

**(b) The calibration layer**

Because the system uses a variety of detection methods, and different methods may generate different output formats. So the output of these different agents must be integrated into a unified format for upper layers to fuse.

**(c). The information layer**

Different basic detection agents may make different judgment from different angle to a certain intrusion, for example: for an U2R attack, the Anomaly Detection agent using state transition technology may generate alarms according to the illegal state transition of a certain user, and it just knows that the some anomalies exist in the system , but doesn't know these anomalies are occurred from which intrusion behavior; A rule-based misuse agent may detect that a U2R intrusion happens in the system correctly; A SVM based classification detection engine may mistake it for a R2L intrusion because there is not enough obvious features; And network-based intrusion agent don’t generate alarms because it cannot detect the intrusion behavior. The most important task of the information layer is to give a reasonable and effective evaluation to the judgments of different kinds of agents to the same event, and finally give a correct decision-making.

**(d) The knowledge layer**

Although the information layer fusion consumedly reduced the amount of original alarms, and the correctness of the decision has been greatly improved than the original decision-making, but the number of alarms is still too much for the system’s administrator, and the decision is still on a relatively low level. So, the result of the information layer still need to be further extracted, so that the administrator can not only acquire system’s security situation from a higher level, but also can make the intrusion scene clear. Therefore, a knowledge layer is intercalated into the system to process the information layer’s output, and to get a more precise and comprehensive understanding to the system’s security situation

**C. Conceptual Model**

Based on analyzing research situation related to network security situation awareness in and abroad, combining analysis of classic models of situation awareness in other domains like JDL functional model[4] and situation awareness mental model proposed by Endsley[5], we give a conceptual model of network security situation awareness. In Fig. , the model consists of three levels, from bottom to top are network security situation perception, situation evaluation, and situation prediction. Level 1. Situation perception is the basis of situation awareness.. This level mainly adopts mature technologies to perceive network security situation information from mass data, translates them into understandable formats (such as XML), and prepares for situation evaluation. Level 2. Situation evaluation is the core of situation awareness, and it is also a dynamic comprehension process of current security situation. It represents security situation of the whole network by recognizing security events, ensuring relationships among events, and generating security situation maps from threats faced. Level 3. Situation prediction is to judge what the future security tendency is, according to the past network security situation and current network security situation. This will help decision-makers know network security situation in a higher level, and supply evidences for reasonable and precise decisions

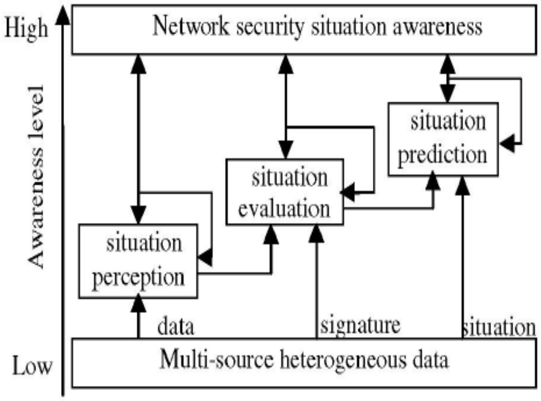

**Figure. 5: Multi Source Heterogeneous Data**

**D. Network security situation awareness Model**

Endsley described SA in three hierarchical phases, as depicted in Figure [2]

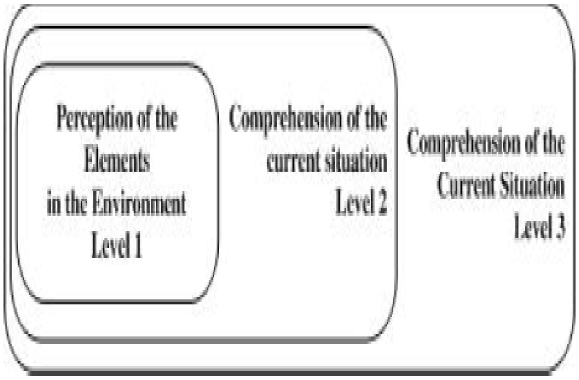
Perception of the Elements in the Environment Level 1
Comprehension of the current situation Level 2
Comprehension of the Current Situation Level 3

**Figure. 6: Level**

Destination we form a ring route such that the two Bass presented a heterogeneous sensors data fusion model to introduce the next generation distributed IDSs [3]. However there are different strengths and weaknesses in different environments and we cannot find a universally accepted one until now and apply it to the application model of NSSA. The model is the basis of NSSA, so we designed one to deploy on our NSSA system according to the model presented in [3] and [6] and it is demonstrated in Figure 7.

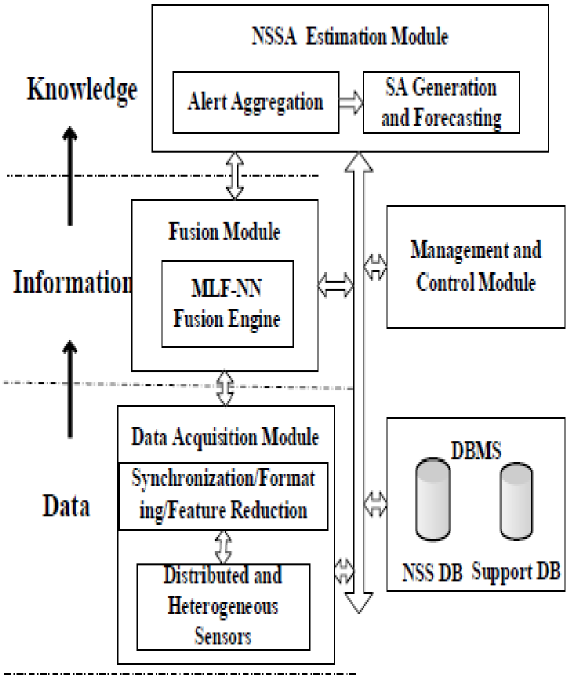

**Figure. 7: Estimation Module**

The first stage of NSSA is Data Acquisition (DA). As a NSSA system, it has three critical characteristics: multi-source, heterogeneity and real-time. What is called multi-source is the data coming from multiple sensors. In our pilot study, we choose Snort and NetFlow as our two heterogeneous sensors to monitor network traffic that result in a cheaper deployment. In the DA module, it also has the ability to wipe off invalidation sensors, format and synchronize the data and extract critical security elements through preprocessing, if necessary. The fusion module fuses the data provided by the DA module and transmits the information obtained from the fusion module to the NSSA Estimation (NSSAE) module. The NSSAE module and the Management and Control (MC) module process the knowledge offered by the fusion module according to the approach that we will discuss below.

## 5. Conclusion

Network security situation awareness system and intrusion detection is a new research domain, and it has great importance in improving abilities of responding to emergences, reducing losses of network attacks, revealing abnormally intrusions, enhancing system abilities of fighting back. The paper gives the study of network security situation awareness model and intrusion detection model. In future we have design a hybrid model combined with data fusion, neural network and D-S Theory.

## References

[1] J.R. Goodall, W.G. Lutters, and K. Anita, "The work of intrusion detection: rethinking the role of security analysts," Proceeding of the Tenth Americas Conf. on Information System, New York, August 2004, pp. 1421-1427.

[2] M.R. Endsley, "Design and Evaluation for Situation Awareness Enhancement," Proceeding of the human factors society 32nd annual meeting, Santa Monica, CA, 1988, pp. 97-101.

[3] T. Bass, "Multi-sensor Data Fusion for Next Generation Distributed Intrusion Detection Systems,"Proceeding of the IRIS national symposium on sensor and data fusion, June, 1999, pp. 99-105.

[4] W. Yurcik, "Visualizing NetFlows for Security at Line Speed: The SIFT Tool Suit." Proceedings of 19th Usenix Large Installation System Administration Conference (LISA), San Diego, CA, USA, Dec. 2005,pp. 169-176.

[5] Carnegie Mellon's SEI. "System for Internet Level Knowledge (SILK)," Http://silktools.source forge.net, 2005.

[6] A.N. Steinburg, C.L. Bowman, and F.E. White, "Revisions to the JDL Data Fusion Model," Joint NATO/IRIS Conference, Quebec, October, 1998.

[7] D.L. Hall, Mathematical Techniques in Multisensor data Fusion. Bosston: Artech House, 2004.

[8] R.Y. Cui, and B.R. Hong, "On Constructing Hidden Layer for Three-Layered Feedforward Neural Networks," Journal of Computer Research and Development, Apr. 2004, Vol. 41, No. 4, pp. 524-530.

[9] X.D. Zhou, and W. Deng, "An Object-Oriented Programming Framework for Designing Multilayer Feedforward Neural Network," Journal of Soochow University, Soochow, China, Feb. 2006, pp. 57-61.

[10] M. Moradi, and M. Zulkernine. "A Neural Network Based System for Intrusion Detection and Classification of Attacks," Proeeding of 2004 IEEE International Conference on Advances in Intelligent Systems, Luxembourg, 2004.

[11] J. Chen, Multisensor management and information fusion. Northwest Industry University,Xian, 2002.

[12] Lincoln Laboratory, Massachusetts Institute of Technology, Darpa Intrusion Detection Evaluation, 2001, Software, Available: http://www.ll.mit.edu 358.

[13] M. Zhang, and J.T. Yao, "A Rough Sets Based Approach to Feature Selection," Proeeding of the 23rd International Conference of NAFIPS, Banff, 2004, pp. 434-439.

[14] R.P. Lippmann, and R.K. Cunningham, "Improving Intrusion Detection Performance Using Keyword Selection and Neural Networks," Computer Networks, 2000, pp. 597-603.

[15] C. Siaterlis, and B. Maglaris, "Towards multisensor data fusion for DoS detection," Proeeding of the 2004 ACM Symp. on Applied Computing, New York, 2004, pp. 439-446.

[16] J.W. Zhuge, D.W. Wang, Y. Chen, Z.Y. Ye, and W. Zou, "A Netwrok Anomaly Detection Based on the D-S Evidence Theory," Journal of Softwoare, March 2006, pp. 463-471.

[17] X.W. Liu, H.Q. Wang, Y. Liang, and J.B. Lai, "Heterogeneous Multisensor Data Fusion with Neural Network: Creating Network Security Situation Awareness," Proceeding of ICAIA, Hong Kong, March 2007, pp. 42-4.

[18] J. Kong, "Anonymous and untraceable communications in mobile wireless networks," Ph.D. dissertation, 2004, chair-Gerla, Mario.

[19] S. network Technologies (SNT) Qualnet, http://www.qualnet.com.