# Intrusion Awareness Based On D-SA

Prashant Shakya[1], Prof. Rahul Shukla[2]
[1,2]Computer Science and Engineering
[1,2]College of Science and Engineering, Jhansi

***ABSTRACT:-***
*Intrusion awareness system is device or software applications that monitor network or system activities for malicious activities or policy violation. Mainly Two types of Intrusion detection systems are network based and host based. This paper is only discussed about network based intrusion system. Matching algorithms are used for detection intrusion. In this paper we proposed a novel method for intrusion awareness using Distributed situational awareness (D-SA). Data fusion work on the biases of features gathering of event. Support vector machine is a super classifier of data. In this paper we used SVM for the detection of closed item of ruled based technique. In Our proposed method, we used KDD1999 DARPA data set and get better empirical evaluation result in comparison of Rule based technique and Distributed Situational Awareness.*

***KEYWORDS****: - intrusion awareness, D-SA, data fusion, SVM and KDDCUP1999.*

## I. INTRODUCTION

Network based intrusion detection system monitor network activities. A network consists of two or more computers that are linked to share resources, exchange files and allow electronic communications. Intrusion detection is the process of monitoring the events occurring in a computer system or network and analyzing them for signs of possible incidents, which are violations or imminent threats of violation of computer security policies, acceptable use policies[9]. Intrusion detection systems (IDPS) are primarily focused on identifying possible incidents, logging information about IDS, and reporting them to security administrators. IDSs basically record information related to observed events, notify security administrators of important observed events, and produce reports. 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 [10]. The DSA methodology comprises three main parts. In the first part, the knowledge owned by each party in each phase of the operation is elicited. Critical Decision Method has been used for this task. The second part is to extract 'knowledge objects' from the Critical Decision Method. Content Analysis has been used for this task. The third and final part is to represent the relations between 'knowledge objects' and identify in which phase(s) they are activated. Propositional Networks were used for this task, comprising 'subject', 'relation' and 'object' network structures of the knowledge required by the system to describe any given situation. In Section II, we present KDDCUP'99 dataset. The Preliminary work of security attack detection and classification is formulated in Section III. In section IV FSVM is proposed. In section V Experimental and result analysis. In section V conclusion and future work.

## II. KDDCUP99 DATASET

To check performance of the proposed algorithm for distributed cyber attack detection and classification, we can evaluate it practically using KDD'99 intrusion detection datasets [6]. In KDD99 dataset these four attack classes (DoS, U2R, R2L, and probe) are divided into 22 different attack classes that tabulated in Table I. The 1999 KDD datasets are divided into two parts: the training dataset and the testing dataset[14]. The testing dataset contains not only known attacks from the training data but also unknown attacks. Since 1999, KDD'99 has been the most wildly used data set for the evaluation of anomaly detection methods. This data set is prepared by Stolfo et al. [11] and is built based on the data captured in DARPA'98 IDS evaluation program [12]. DARPA'98 is about 4 gigabytes of compressed raw (binary) tcp dump data of 7 weeks of network traffic, which can be processed into about 5 million connection records, each with about 100 bytes. For each TCP/IP connection, 41 various quantitative (continuous data type) and qualitative (discrete data type) features were extracted among the 41 features, 34 features (numeric) and 7 features (symbolic). To analysis the different results, there are standard metrics that have been developed for evaluating network intrusion detections. Detection Rate (DR) and false alarm rate are the two most famous metrics that have already been used [16]

**TABLE I. DIFFERENT TYPES OF ATTACKS IN KDD99 DATASET**

| 4 Main Attack Classes | 22 Attack Classes |
|---|---|
| Denial of Service (DoS) | back, land, neptune, pod, smurt, teardrop |
| Remote to Local (R2L) | ftp_write,guess_passwd, warezclient,warezmaster |
| User to Root (U2R) | buffer_overflow,perl, |
| Probing(Information Gathering) | ipsweep, nmap, portsweep, satan |

DR is computed as the ratio between the number of correctly detected attacks and the total number of attacks, while false alarm (false positive) rate is computed as the ratio between the number of normal connections that is incorrectly misclassified as attacks and the total number of normal connections. In the KDD Cup 99, the criteria used for evaluation of the participant entries is the Cost Per Test (CPT) computed using the confusion matrix and a given cost matrix .A Confusion Matrix (CM) is a square matrix in which each column corresponds to the predicted class, while rows correspond to the actual classes. An entry at row i and column j, CM (i, j), represents the number of misclassified instances that originally belong to class i, although incorrectly identified as a member of class j. The entries of the primary diagonal, CM (i, i), stand for the number of properly detected instances. Cost matrix is similarly defined, as well, and entry C (i, j) represents the cost penalty for misclassifying an instance belonging to class i into class j. Cost matrix values employed for the KDD Cup 99 classifier learning contest are shown in Table 2. A Cost Per Test (CPT) is calculated by using the following formula:[17]

$$CPT = 1/N \sum_{i=1}^{m} \quad \sum_{j=1}^{m} CM(i,j) * C(i,j)$$

Where CM and C is confusion matrix and cost matrix, respectively, and N represents the total number of test instances, m is the number of the classes in classification. The accuracy is based on the Percentage of Successful Prediction (PSP) on the test data set.

$$\mathbf{PSP} = \frac{\textbf{number of successful instance classification}}{\textbf{number of instance in the test set}}$$

## III. PROPOSED METHOD

**DSA Methodology**

Distributed Situation awareness is the perception of environmental elements with respect to time and/or space, the comprehension of their meaning, and the projection of their status after some variable has changed, such as time, or some other variable, like a predetermined event. It is a field of study concerned with perception of the environment critical to decision-makers in complex**.** There are some steps of distributed situational awareness

1. Elicit the knowledge owned by network
2. Extract knowledge objects
3. Shows the relations between knowledge objects and their activation
4. Compute value of percentage of successful prediction
5. Compute value of cost effective per text
6. Compute value of detection rate

The methods indicate that there is a lot of teamwork occurring in each scenario although there is a clear hierarchy. The PWO and AAWO will remain the central nodes of the operations room. Information is shared between the crew members Shared awareness can be seen from the analysis in table two. Knowledge objects are shared within the three individual scenarios (i.e., air, surface and sub-surface) as well as across the whole mission. It is important to remember that the three scenarios are observed and often happen at the same time and will not be separated into three clear areas. Thus the sharing of knowledge objects across different scenarios will be essential for effective operations.

## IV. EXPERIMENTAL RESULTS

All the experiments were performed on an Intel ® Core ™ i3 with a 2.27GHz CPU and 4 GB of RAM. We used MATLAB version 2013 software. To evaluate the performance of our proposed cyber attack detection system, we used the KDDCUP1999 dataset. Our experiment is split into three main steps. In the first steps, we prepare different dataset for training and testing. Second, we apply distributed situational awareness algorithm (DSA) to the dataset. The original KDDCUP1999 dataset to select most discriminate features for intrusion attack detection. Third, we classify the intrusion attacks by using rule based as classifier. For the performance evaluation we used two different data set of KDDCUP99.

**Table 2.Comparative result of rule based method and DSA**

| Data set | Method | Detection rate |
|---|---|---|
| Data set1 | Rule classification | 92.6231 |
| | DSA | 95.6321 |
| Data set2 | Rule classification | 93.231 |
| | DSA | 96.6231 |

## V. CONCLUSION

In this paper we proposed a new method for security alert generation for intrusion awareness. Such method based on distributed situational awareness. This approach can discover new alert relations and does not depend on background knowledge. At last, we tested our methods on DARPA 2000 Dataset. The simulations showed that with the proposed methods DSA system can efficiently analyze large amount alerts and save administrators' time and energy. In future we used auto correlation for better prediction of precision and recall.

## 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'07, 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