Novel Technique of Extraction of Principal Situational Factors for NSSA

Pardeep Bhandari,
Asst. Prof.,Computer Sc., Doaba College, Jalandhar
bhandaridcj@gmail.com

Abstract— *The research on Network Security Situational Awareness has become hot area because of increase in reliance on computer networks. The variety of services being provided on the networks has increased many folds. Major problem in this field is to perceive the security situation of the network because of large volume of data produced per unit time, even in a moderate size network. Though the computing capacities of modern machines have increased but to perceive the security situation, very heavy real time data is to processed, which has become a challenge even for modern computing facilities. In this paper data preprocessing technique based feature selection has been proposed. Features reduction is performed using chisquare attribute evaluation and ranker search method. To ascertain the classification performance using reduced feature set Bayesnet and Naivebayes classifiers are used. Current study uses KDD Cup 1999 Train+ data sets as experimental data and comes to conclusion that better situation perception may be achieved by using a small subset of the attributes of dataset. The members of the selected dataset may then be used as situational factors for further analysis of security situation.*

Index Terms—*Bayesnet classification algorithm, feature selection, situational awareness, situation prediction.*

I. INTRODUCTION

According to[1] the term “situation Aawareness” has been often described as merely the knowledge of all the objects in a specific area, and possibly their relative states. It is clear, however, that interpretation of this term implies much more than that. For instance, in Merriam-Webster Online (2004), “AWARE implies vigilance in observing or alertness in drawing inferences from what one experiences.” The next question is about the meaning of being aware of a specific situation. The word “situation” is defined in Merriam-Webster Online (2004) as: “the way in which something is placed in relation to its surroundings.” So the essence of situation awareness [2] is the knowledge of all the individual elements related to the subject under consideration, the knowledge of static and dynamic relations among the elements, and elements and subject. Static relations may be hierarchical in nature and dynamic relation exists when change in characteristic of one element affects other element. These individual elements are termed as situational factors. Working on these lines, in case of computer networks, the network connections established among the nodes must be considered to be one of the situational factors to perceive the current security situation. There are so many other factors like users, user privileges, installed software & hardware, vulnerabilities detected, topology etc. TCP connection is considered to be the most important of all the situational factors, because any type of attack in a computer network is performed using TCP connections. To describe one network connection, 41 attributes have been considered in KDD Cup 1999 dataset. Reduction in feature vector is of utmost importance for NSSA. The volume of data generated and to be processed to perceive the security situation of a network is very large. Processing of this data in real time is very difficult. Late retrieval of information may result in missing the security alert. Therefore, in this paper we are proposing improved technique to identify most important attributes or in other words, reduce feature vector of TCP connection.

II. LITERATURE REVIEW

In 1999, Tim Bass[3] first proposed the concept of cyberspace situation awareness based on Endsley's[4] and established a functional framework for it, which provided a theoretical foundation for research on *Network Security Situational Awareness* **(**NSSA). Tim bass has proposed the concept of *NSSA*, which considers the network security holistically. NSSA is defined as a system that allows network security manager to understand and evaluate the network security holistically. The realization of NSSA is divided into three layers[5][6][7]. First is perception of Situational Factors i.e. Situation Perception. Second is comprehension of situation factors (SF), which involve combination and explanation of SFs. The third layer is projection or situation prediction which deals with forecast of the future network security situation. Aiping Lu et. al. [8] has proposed data preprocessing technique based on conditional random field. But they have validated their results on different categories of attacks, so the technique proposed performs well in identifying only some specific categories of attacks. Huiqang Wang et.al.[9]have proposed method to extract situational factors based on evolutionary neural network. They have compared the extraction capabilities of evolutionary strategy based neural network and genetic algorithm based neural network. Bhandari A. [10] has proposed technique for intrusion detection system that uses rough set theory for feature selection, which is extraction of relevant attributes from the entire set of attributes describing a data packet and used the same theory to classify the packet if it is normal or part of an attack on the network. Xiaowu Liu et. al. [11] have worked on multi sensor data fusion and feature reduction using support vector machine and have reported good results. The efficiency of these proposed techniques may further be improved by improvement in process of feature reduction. This has been the motivation for this paper, which proposes a method for first layer i.e. situation perception. Data produced per unit time in a network of moderate size is extremely large. NSSA has to predict the future situation of the network based on the past security situation as well as the current network condition in shortest possible period of time. If the data is not preprocessed properly at first layer, it may affect the situation evaluation and situation prediction i.e second and third layer of NSSA. In this paper we propose a technique based on Bayesian Network to identify the most important characteristics to describe current situation of the network. The paper has been organized as follows: In section III Classification and feature selection problems are introduce. Two classifiers i.e. bayesnet classifier and naivebayes classifier are introduced. For feature selection, attribute evaluators and various search methods have been introduced. Section IV gives description of feature set used in the dataset used for experimentation. Section V gives the proposed methodology for data preprocessing and also discusses the evaluating indexes used to compare the performance of attribute evaluator and search methods. Section VI discusses results and analysis and conclusion and future work is given by section VII.

III. CLASSIFICATION PROBLEM

Classification is a process of classifying the objects on the basis of fixed characteristics using a model. The model is built using known values, also called input variables [17]. This model is then used to predict unknown values, termed as output variables. The model used for classification is developed to represent the relationship between the input variables and the output variables which is to be predicted. The process of model development required data in which both the input variables and the output variables are present. This step when initial model is presented with labeled data is known as training of model. This model can subsequently be used on unseen data in which only the input data is present, the output is performed by model on the basis of its learning. The process of classification is categorized as supervised learning because it is trained specially for the purpose of classification. Classification is a basic task in data analysis and pattern recognition that requires the construction of a *classifier*, that is, a function that assigns a *class* label to instances described by a set of *attributes*. In the current paper we have used two classifiers namely naive Bayesian classifier and Bayesnet classifier [12][13][14][15]. Naïve Bayes classifier is based on Baye's theorem and assumes strong independence assumptions between the features.Naïve Bayes classifiers learns from training data, the conditional probability of each attribute A_i given the class label C. Classification is then done by applying Bayes rule to compute the probability of C given the particular instance of $A_1,\ldots,$ A_n, and then predicting the class with the highest posterior probability. This computation is rendered feasible by making a strong independence assumption: all the attributes $A_i$are conditionally independent given the value of the class C. Probabilistic independence implies that A is independent of B given C whenever

$$P(A/B,C) = P(A/$$

For all possible values of A, B and C, whenever P(C) >0 where P stands for probability

The performance of Naive Bayes is somewhat surprising, since the above assumption is clearly unrealistic in most situations. In case of computer networks attributes describing a TCP connections are related to each other in more than one ways[16].Bayesnet classifier does not assume independence between the features. Suitability of the classifier is based on the predictive accuracy i.e. ability of the model to correctly ascertain the class of previously unseen data, accuracy i.e. percentage of testing set examples correctly classified by the classifier, speed, robustness, scalability etc. In this paper we have classified the KDD Train+ test data consisting of 41 attributes and 125973 instances using bayesnet and naivebayes classifier and compared their performance. The performance of both the classifiers has been compared on the basis of various parameters shown on Table-I. The results i.e. value of accuracy, precision, recall rate, F-Value, Detection Rate show that classification accuracy of Bayesnet classifier is much better in all respects. Also False Alarm rate, Missing Alarm rate and Overall error are least in case of Bayesnet classifier.

TABLE-I

COMPARISON OF CLASSIFIERS ON EVALU

CLASSIFIER TYPE	NAIVEBAYES
ACCURACY	0.905932619
PRECISION	0.924658994
RECALL RATE	0.869047023
F-VALUE	0.895990913
DETECTION RATE	0.869047023
FALSE ALARM RATE	0.061849357
MISSING ALARM RATE	0.130952977
OVERALL ERROR	0.094067381

Feature selection is a process of identifying the smallest number of attributes to describe an entity. The objective of feature selection in machine learning is to improve accuracy and reduce training time for the model.The ability to apply feature selection is critical for effective analysis, because datasets frequently contain far more information than is needed to build the model. Even if resources like memory and processing time are not an issue, you typically want to remove unneeded columns because they might degrade the quality of discovered patterns[17][18]In this paper we are going to use Weka Explore 3.6.1[22] for feature selection. Basically the process of feature reduction consists of two steps. In first step space of possible subsets of attributes is searched and in second step the identified subset is assessed. This process is repeated and the most suitable subset is given as the result.

The Search Method is the planned way in which the search space of possible attribute subsets is navigated based on the subset assessment or evaluation. Search may be either Random Search or Exhaustive Search. Some examples of attribute evaluation methods are: Exhaustive search, which tests all combinations of attributes. Best-First search uses a best-first search strategy to navigate attribute subsets. GreedyStepWise uses a forward (additive) or backward (subtractive) step-wise strategy to navigate attributes subsets. There are broadly two categories of feature selection algorithms namely wrapper and filters. The wrapper category is further divide into forward greedy wrapping i.e. keep adding features one at a time until no further improvement can be achieved and backward greedy wrapping i.e. keep removing features one at a time until no further improvement can be achieved. A third alternative is to interleave the two phases (adding and removing) either in forward or backward wrapping ("forward-backward wrapping") [17]. In the filter approach we do not rely on running a particular classifier and searching in the space of feature subsets; instead we select features on the basis of statistical properties.

Some examples of attribute evaluation methods are

CfsSubsetEval : this method gives priority to those subsets that correlate highly with the class value and low correlation with each other,*ClassifierSubsetEval*: this method assesses subsets using a predictive algorithm and another dataset specified by the user, *WrapperSubsetEval*: assesses subsets using a classifier that you specify and n-fold cross validation is carried out [20][21].

If A_1, A_2 ….A_k represents the subsets of attribute set S. The subsets are identified based upon some searching algorithm discusses above. Performance of each subset is compared with performance of sample independent best feature set A_{best} in terms of error $ἐ_k$

of subset $A_k$and ξ_{best} error with A_{best}[19]. The subset for which $ἐ_k \approx \xi_{best}$is selected as reduced feature set.

IV. DESCRIPTION OF FEATURE SET

Experimental data used in this paper is KDD cup 1999 data sets[23] from standard database. We have specifically used KDD Train+ dataset, which contain marked instances of normal and anomalous network flow i.e. TCP connection. A connection is a sequence of TCP packets starting and ending at some well defined times, between which data flows to and from a source IP address to a target IP address under some well defined protocol. Each connection is labeled as either normal, or as an attack, with exactly one specific attack type. For each instance there are 41 attributes, which can be classified as three features sets namely *basic feature set, content feature set, flow and traffic of hosts feature set*. *Basic feature set* consists of duration, protocol_type, service, src_bytes ,dst_bytes, flag, land , wrong_fragment , and urgent. *Content feature set* consists of hot, num_failed_logins ,logged_in, num_compromised, root_shell, su_attempted, num_root , num_file_ creations,num_shells,num_access_files, num_ outbound_cmds, is_hot_login,andis_guest_ log_in. *Flow and traffic feature set* consist of count, serror_rate,rerror_rate, same_srv_rate, diff_srv_rate, srv_count, srv_serror_rate, srv_rerror_rate, and srv_diff_host_rate.

V. Methodology for Data Preprocessing

Steps of methodology are as follows:

- The dataset is classified based on naïvebayes and bayesnet classifier using all the 41 attributes. 66% of the dataset is used as training data, which is used to develop the model; rest of the data is used as test data.
- Classification results have been compared based on the evaluation indexes discussed above. Bayesnet has proved to be better choice, so the classification results given by bayesnet have been considered as standard.
- Features are ranked using chi square attribute evaluator and Ranker search method.
- From the ranked set of features, classification is repeated using 11, 21, 31 and 41 number of attributes using bayesnet classifier. The attributes are picked as per the ranking provided by the previous step.
- Classification results of each case are compared. This comparison gives the best subset of attributes which is capable of performing classification leading to lower computing cost and accuracy.

To compare the classification accuracy of classifiers and classification accuracy of classifiers based on different types and number of attributes, following evaluating indexes have been used[8]. Positive sample shows presence of phenomenon of interest and negative sample shows it absence. For instance, if the data consists records representing normal state of network and under attack state of network, then presence of attack is termed as positive sample and normal state is termed as negative sample:

- Accuracy= (TP+TN)/ (P+N)
- Precision or Positive Prediction Value= TP/(TP+FP)
- Recall rate =TP/(TP+FN) an estimate of the probability that a positive observation will be classified as positive.TPR
- F-Value = (2*precision*recall rate)/(precision+recall rate), used in case when no. of negative cases is much greater than the number of positive cases(Kubat et al., 1998).
- Detection Rate= TN/TN+FP i.e. number of correctly detected intrusions i.e. anomalies/ total no. of intrusions
- False Alarm Rate=FN/TP+FN
- Missing Alarm Rate=FP/TN+FP

Total no. of samples=P+N

Where P, N: no. of positive & negative samples respectively

TP, TN: no. of correctly classified positive samples & negative samples respectively

FP=No. of negative sample incorrectly classified as positive

FN= No. of positive sample incorrectly classified as negative.

VI. Results and Analysis

Classification of dataset based on naïvebayes classifier and bayesnet classifier yield Confusion Matrices given in Fig.-1 and Fig.-2

=== Confusion Matrix ===		
a b ← classified as		
21448	1414	a = normal
2615	17354	b = anomaly

Fig.-1 Confusion Matrix Naivebayes Classifier

=== Confusion Matrix ===		
a b ← classified as		
22715	147	a = normal
1007	18962	b = anomaly

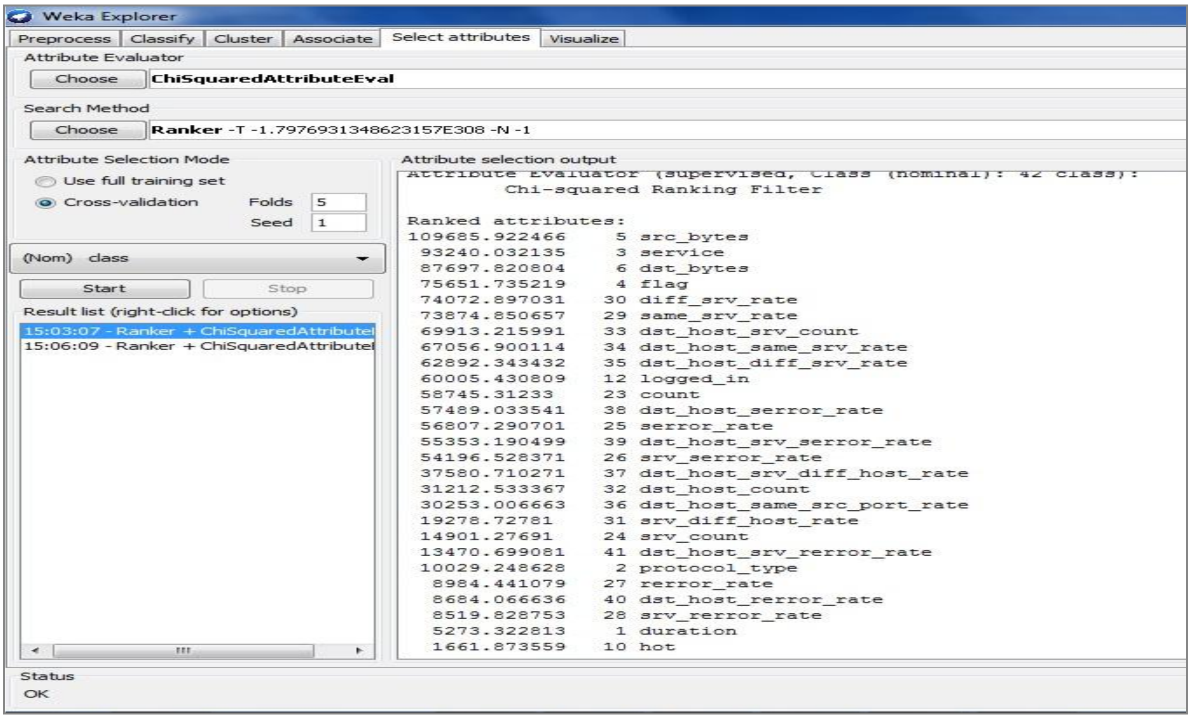
Fig.-2 Confusion MatrixBayesNet Classifier

In Fig.I and Fig. 2 confusion matrices, it is clear that True Positives in case of NaiveBayes classifier(21448) is lesser than True Positives in case of BayesNet(22715) i.e. BayesNet has identified normal situation of network more precisely. Also Bayesnet has identified only 147 negative instances as positives i.e. False Positive in comparison to

NaiveBayes whose number of False Positives is 1414. Similar results are there for anomalous situation of the network.

Table-I shows that classification based on bayesnet classifier outperforms the classification based on naivebayes. This is because of independence assumption on which naivebayes works. Therefore results given by bayesnet classifier have been considered as standard to check the performance of classification based on different subsets of the attributes from the dataset.

To rank the features of the dataset ranker search method is used to search subsets of the attributes and chi square attribute evaluation method is used for evaluation. Following ranking is obtained as shown in Fig-3:



Attribute Evaluator: ChiSquaredAttributeEval

Search Method: Ranker -T -1.7976931348623157E308 -N -1

Attribute Selection Mode
○ Use full training set
● Cross-validation   Folds 5   Seed 1

(Nom) class

Result list (right-click for options)
15:03:07 - Ranker + ChiSquaredAttributeI
15:06:09 - Ranker + ChiSquaredAttributeI

Attribute selection output

```
Attribute Evaluator (supervised, Class (nominal): 42 class):
        Chi-squared Ranking Filter

Ranked attributes:
109685.922466    5 src_bytes
 93240.032135    3 service
 87697.820804    6 dst_bytes
 75651.735219    4 flag
 74072.897031   30 diff_srv_rate
 73874.850657   29 same_srv_rate
 69913.215991   33 dst_host_srv_count
 67056.900114   34 dst_host_same_srv_rate
 62892.343432   35 dst_host_diff_srv_rate
 60005.430809   12 logged_in
 58745.31233    23 count
 57489.033541   38 dst_host_serror_rate
 56807.290701   25 serror_rate
 55353.190499   39 dst_host_srv_serror_rate
 54196.528371   26 srv_serror_rate
 37580.710271   37 dst_host_srv_diff_host_rate
 31212.533367   32 dst_host_count
 30253.006663   36 dst_host_same_src_port_rate
 19278.72781    31 srv_diff_host_rate
 14901.27691    24 srv_count
 13470.699081   41 dst_host_srv_rerror_rate
 10029.248628    2 protocol_type
  8984.441079   27 rerror_rate
  8684.066636   40 dst_host_rerror_rate
  8519.828753   28 srv_rerror_rate
  5273.322813    1 duration
  1661.873559   10 hot
```

Fig. 3: Snapshot of Ranking of attributes in WEKA

Selected attributes are as follows: 5,3,6,4,30,29,33,34,35,12,23,38,25,39,26,37,32,36,31,24,41,2,27,40,28,1,10,8,13,16,19,22,17,15,14,18,11,7,21,20,9 : 41

In Case 1: Total attributes=11, First 10 attributes (attributed no. 5,3,6, 4,30,29,33,34,35,12,23,41) along with class representing normal or anomalous condition are considered. Classification is performed using bayesnet classifier using these 11 attributes

Case 2: Total attributes=21, First 20 attributes (5,3,6,4,30,29,33,34,35,12,23,38,25,39,26,37,32,36,31,24,41) along with class representing normal or anomalous condition are considered. Classification is performed using bayesnet classifier using these 21 attributes

Case 3: Total attributes=31 First 30 attributes(5,3,6,4,30,29,33,34,35,12,23,38,25,39,26,37,32,36,31,24,2,27,40,28,1,10,8,13,16,19,41) along with class representing normal or anomalous condition are considered. Classification is performed using bayesnet classifier using these 31 attributes.

TABLE –II
VALUES OF CONFUSION MATRICES OBTAINED IN CASE 1,2,3,4

	Case 1: 10 Attributes	Case 2: 20 Attributes	Case 3: 30 Attributes	Case 4: 41 Attributes
TN	22464	22745	22715	22715
FP	398	117	147	147
FN	759	1053	1010	1007

Case 4: Classification is performed using all the 41 attributes with bayesnet classifier. Values of confusion matrices of above four cases are shown in Table-II and the classification results in terms of evaluating indexes are given in Table-III

TABLE III
COMPARISON OF EVALUATING INDEXES IN CASE-1,2,3,4

	CASE 1	CASE 2	CASE 3	CASE 4
ACCURACY	0.973	0.973	0.973	0.973
PRECISION	0.980	0.994	0.992	0.992
RECALL RATE	0.962	0.947	0.949	0.950
F-VALUE	0.971	0.970	0.970	0.970
DETECTION RATE	0.962	0.947	0.949	0.950
FALSE ALARM RATE	0.017	0.005	0.006	0.006
MISSING ALARM RATE	0.038	0.053	0.051	0.050
OVERALL ERROR	0.027	0.027	0.027	0.027

The results show that Accuracy achieved using 10 attributes is equivalent to using 20, 30 or all the attributes of the dataset. Precision achieved using 10 attributes from the ranking is 98% which is quite good. Recall rate 0.962 has been achieved which is more than using larger number of attributes. F-Value measure and detection rates are also greatest in Case 1 i.e. using ten attributes. Similarly False alarm rate, missing alarm rate and overall error achieved are best as compared to classification performed in case2,3,4.

Fig-4 shows that the classification performed using first 10 attributes from the ranking produces optimal results comparable with results obtained using larger number of attributes. This implies that storage and observation of only these top ten features is able to predict the intrusion in the network at any given instance. This observation will definitely lead to saving in computation time and computing resources required to predict current status of the network in case of enterprise networks.

VII. Conclusion

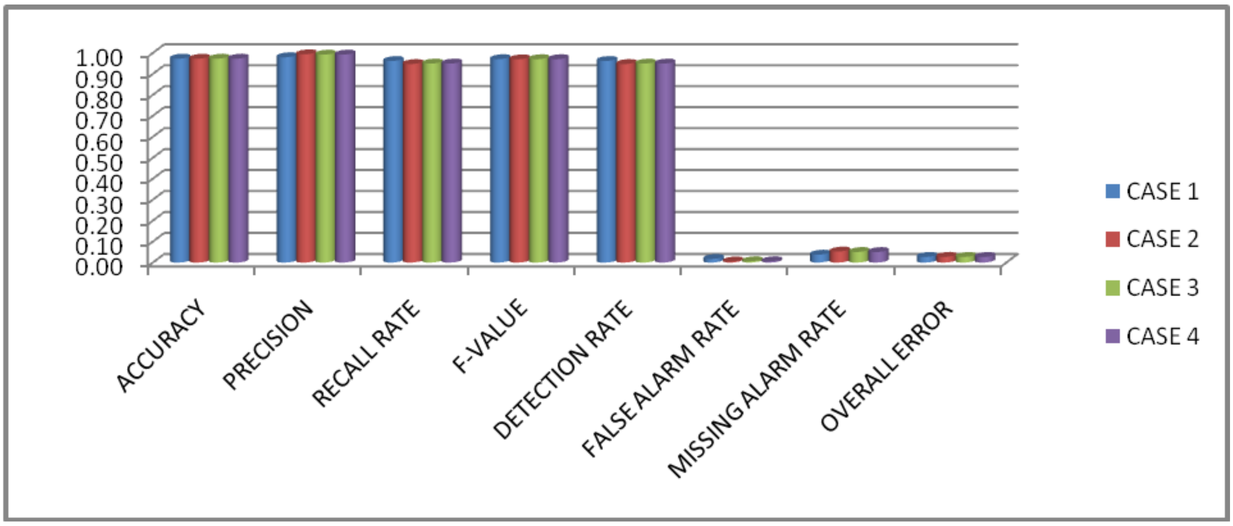


Fig-4 Curves comparing the values of evaluating indexes for Case 1,2,3,4

In this paper we have proved that data preprocessing method based on chi squared attributed evaluator and Ranker search method may provide substantial decrease in data to be handled in network security situational awareness. Also classification based on bayesnet classifier may be used for situation perception based on smaller number of attributes. This is substantial improvement in timely perception of current state of network security in NSSA. The identification of subset of most important attributes guides the network administrator to focus on most important happenings in the network. In case of dataset consisting of billions of records identification of smaller subset of attributes also leads to substantial saving in processing time thus giving the timely information to the network administrator. The feature reduction technique identified in this paper will be helpful in situation comprehension i.e. second layer of NSSA. It will give us the attributes to be used to comprehend the situation and also to predict future state of network security of the network, which is the actual objective of NSSA. This will be our future direction of research.

References

[1] M. M. Kokar and H. Avenue, “Situation Awareness : Issues and Challenges.”

[2] K. Baclawski and C. J. Matheus, “Formalization of Situation Awareness,” pp. 1–14.

[3] T. Bass, “a glimpse into the future of id,”Available at: http://www.usenix.org/publications/login/1999-9/features /future.html,1999.

[4] Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors: The Journal of the Human Factors and Ergonomics Society. doi:10.1518/001872095779049543

[5] J. Han, “Compilation and Evaluation of Linear Mutual Recursions,” Inf. Sci. (Ny)., vol. 69, pp. 157–183, 1993.

[6] J. Wang, Z. Qin, and L. Ye, “Modeling of Network Situation Awareness,” no. 2006, pp. 461–465, 2008.

[7] Z. Yong, T. Xiaobin, and X. Hongsheng, “A Novel Approach to Network Security Situation Awareness Based on Multi-Perspective Analysis,” 2007 Int. Conf. Comput. Intell. Secur. (CIS 2007), pp. 768–772, Dec. 2007.

[8] A. Lu, J. Li, and L. Yang, “A New Method of Data Preprocessing for Network Security

Situational Awareness,” 2010 2nd Int. Work. Database Technol. Appl., pp. 1–4, Nov. 2010.

[9] H. Wang, Y. Liang, and H. Ye, “An Extraction Method of Situational Factors for Network Security Situational Awareness,” 2008 Int. Conf. Internet Comput. Sci. Eng., pp. 317–320, Jan. 2008.

[10] A. S. Bhandari, “FEATURE SELECTION AND CLASSIFICATION OF INTRUSION DETECTION SYSTEM USING ROUGH SET,” no. 2, pp. 20–23, 2013.

[11] X. Liu, H. Wang, J. Lai, Y. Liang, and C. Yang, “Multiclass Support Vector Machines Theory and Its Data Fusion Application in Network Security Situation Awareness,” 2007 Int. Conf. Wirel. Commun. Netw. Mob. Comput., pp. 6343–6346, Sep. 2007.

[12] C. S. Division, M. Park, P. Langley, and P. Smyth, “Bayesian Network Classifiers *,” vol. 163, pp. 131–163, 1997.

[13] N. L. Zhang and D. Poole, “A simple approach to Bayesian network computations,” in Proceedings of the 10th Biennial Canadian Artificial Intelligence Conference, 1994, pp. 171–178.

[14] D. Fierens, H. Blockeel, J. Ramon, and M. Bruynooghe, “Logical Bayesian networks,” in Multi-Relational Data Mining ({MRDM} 2004), 2004, pp. 19–30.

[15] Z. Markov and I. Russell, “Probabilistic Reasoning with Naïve Bayes and Bayesian Networks,” 2007.

[16] P. Iyer and D. S. Reeves, “Reasoning About Complementary Intrusion Evidence,” 20th Annu. Comput. Secur. Appl. Conf., pp. 39–48.

[17] M. A. Hall, “Correlation-based Feature Selection for Machine Learning,” 1999.

[18] Zhenglu Yang, “Applied Data Mining” , CRC Press 2013, pp. 100–116, eBook ISBN: 978-1-4665-8584-3, DOI: 10.1201/b15027-7

[19] Microsoft Developer Network, September 2014 Available at: http://msdn.microsoft.com/en-IN/library/ ms175382.aspx.

[20] Jason Brownlee, Septempber 2014, “Feature Selection to Improve Accuracy and Decrease Training Time” Available:http://machinelearningmastery.com/ feature–selection-toimprove-accuracy-and-decrease-training- time/

[21] Feature Selection, School of Medicine, New York University, Available at: http://webdoc. nyumc. org/ nyumc/files/chibi/attachments/file7.pdf.

[22] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.

[23] http://kdd.ics.uci.edu/databases/kddcup99/kdd cup99.html